Preferred citation: M.B. Lyne, A. Parush, F. Richter, B.D. Jordan, D.C. Donderi and J.O. Ramsay. Multidimensional analysis of paper-related factors in the subjective evaluation of print quality. In **The role of fundamental research in paper-making**, *Trans. of the VIIth Fund. Res. Symp. Cambridge, 1981*, (Fundamental Research Committee, ed.), pp 655–683, FRC, Manchester, 2018. DOI: 10.15376/frc.1981.2.655.

MULTIDIMENSIONAL ANALYSIS OF PAPER-RELATED FACTORS IN THE SUBJECTIVE EVALUATION OF PRINT QUALITY

M.B. Lyne, A. Parush*, F. Richter**, B.D. Jordan, D.C. Donderi*, and J.O. Ramsay*
Pulp and Paper Research Institute of Canada
Pointe Claire, Quebec, Canada
* Department of Psychology, McGill University, Montreal, Canada
** University of California, Los Angeles, California

Abstract

Multidimensional scaling, a statistical technique that permits separation and identification of the principal factors used by people when judging differences and preferences between pairs of test stimuli, has been adapted to the subjective evaluation of print quality. Numerical values of the factors involved in subjective print quality evaluation are used to establish the relationship with corresponding physical print qualities and related paper properties. Information is also generated concerning the preferences and reliability of each judge and the degree to which each judge agrees with other judges in a professional group.

Multidimensional analysis of the subjective evaluation of wire-mark in solid letterpress prints indicates that the degree to which lines appear in the wire-mark pattern is as disturbing as the overall wiremark intensity.

Mottle, show-through, contrast, and paper colour are found to be of importance in the judgement of stereo letterpress print quality; while mottle, linting, show-through, and set-off are found to be significant in polymer plate letterpress printing trials. Physical tests are compared for their ability to predict these subjective print quality factors.

Introduction

The marketability of most paper ultimately depends on subjective evaluation of its end-use qualities. The cleanliness of photographic paper, the tactile softness of tissues and the print mottle of gravure printing paper are examples of subjective properties that are judged by the consumer. The paper industry may have physical or chemical tests which relate to these properties, but in the final analysis it is human perception that counts.

In order to improve consumer acceptance of a product it is first necessary to identify the criteria used in judging its qualities. The relative influence each of these criteria has on the preferences of the consumer must then be assessed to give an indication of the most effective way in which to alter the properties of the product. Since preferences are likely to vary from person to person it may also be useful to have a means of comparing individual preferences with those of the group of consumers to which the individual belongs. A numerical, or metric basis for scaling the quality criteria and their influence on preferences is desirable if the amount by which the properties of the product are to be changed is to be specified.

A metric approach to the assessment of the subjective qualities of the product and related consumer preferences also permits statistical analysis of important questions, such as the confidence that can be placed on the judgements, and the strength of consumer preferences. Furthermore, metric scaling allows hypothesis testing, such as whether certain physical tests are significantly related to the criteria used in subjective evaluation, or whether one physical test is better related than another.

Multidimensional scaling (MDS) is a statistically based metric scaling technique that is particularly suited to this sort of investigation. Founded on the concept that people use several factors when judging the difference between complex items, it is designed to separate and quantify the importance of these factors in human judgement. For example, it has been used to map the factors involved in the subjective response to differences in perceived colour(1); to identify and scale the character traits employed by subjects in the judgement of their relationship to others(2); and in the study of how various characteristics of noise interfere with the visual perception of form in images(3). Lyne(4) has described how MDS could be used to identify and scale the factors used by groups of paper-makers, printers, advertisers and readers in the judgement of the quality of images printed on papers having different physical properties. In this article the application of MDS to the evaluation of print quality is extended to an analysis of the relationships between the subjective factors used in the judgements and the physical properties of the substrate papers. While the cases cited in this article deal with the print quality of newsprint the techniques described are equally applicable to the analysis of other subjectively judged attributes of paper.

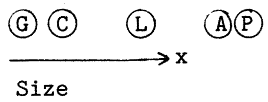
Multidimensional scaling

Concept

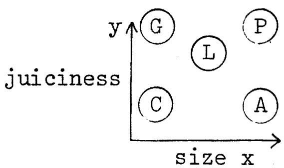
Torgerson(5,6) established as the basic principle of MDS that perceived differences between items can be represented by physical distances, and that these distances can be used to construct a configurational model of the relationship among the items. This process can be illustrated by asking a subject to mark, on a scale from one to ten, the differences perceived between pairs of the following five fruit: cherry, grape, lemon, apple, peach. Let us suppose the responses for the ten possible pairs were :

C	5			
G	5	4		
A	5	5	7	
P	5	7	5	4
	L	C	G	A

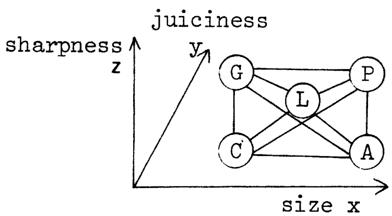
If these perceived differences were simply equated to distances between the fruit, and the fruit were spread out along a line to best fit the measurements, the configuration would be something like this:



Examination of this configuration suggests that the subject had judged the differences between the fruit according to size. Since the fit is imperfect a two dimensional model could be tried and the configuration would be as below:



This configuration suggests that the subject judged the fruit according to juiciness as well as size, but that the differentiation was greater for size. A three dimensional model fits the data best and would form a slightly distorted pyramid.



It is apparent that while the lemon is midway in size and juiciness, it is distinct from the other fruit in having a much sharper taste. The size, juiciness and sharpness can be numerically scaled using the distance information so that each fruit can be assigned co-ordinates on these subjective axes.

Mathematical Basis and Assumptions

The simplest manner in which to convert a set of differences between items into a multidimensional configuration is to apply the generalised Pythagorean theorem.

$$d^*_{ij} = \sqrt{\sum_m (X_{im} - X_{jm})^2} \tag{1}$$

where

d^*_{ij} = model predicted difference between items i and j

X_{im}, X_{jm} = co-ordinates of items i and j, respectively, in the m^{th} dimension.

In fact, Shepard[7,8] has shown that a configurationally consistent model of the items can be created by fitting co-ordinates using any monotonic mathematical function (a function that preserves the ordering of a set of samples when used to transform differences into multidimensional distances).

Since it is desired to compare the responses of various judges and ultimately to combine their responses into a group configurational model it is necessary to adjust for differences in the styles of response. In the example, the judgements of the difference between pairs of fruit were clustered in the centre of the scale, whereas, generally, there is some tendency for judgements to cluster towards the ends due to the ceiling effect of a finite scale. In order to match the response characteristics of different judges, Ramsay[9] added two parameters, V and p, to the basic Euclidean model (Eq. 1) which act as scalar and power corrections, respectively, without altering the monotonicity of the difference-distance relationship:

$$d^*_{ij} = V\left[\left(\sqrt{\sum_m (X_{im} - X_{jm})^2}\right)\right]^p \qquad (2)$$

For each judge, this model is fitted to the perceived differences for all the pairs of items ij, by altering the model parameters in iterative steps in order to maximise the log likelihood(10). The log likelihood can be calculated as follows:

$$\text{Log } L = -M(1 + \log s^2)/2 \qquad (3)$$

where M = total number of judged pairs ij
and s = maximum likelihood estimator for the standard error.

Since the error in judging the difference between two items varies proportionally with the magnitude of that difference (people are more precise in judging differences between similar items than disparate items), the error in making replicate difference judgements is assumed to be distributed about the model predicted value of difference d^* with a lognormal distribution, $N(\log d^*, \sigma^2)$, where σ is the standard error of the log of the observed differences.

An unbiased estimate for this standard error can be calculated as follows:

$$s^2 = \sum_{ij} \frac{(d_{ij} - d^*_{ij})^2}{(M-Q)} \qquad (4)$$

where Q = number of parameters in the model,
d_{ij} = observed difference between items i and j,
d^*_{ij} = model predicted difference between items i and j.

The optimisation of the fit of the model to the observed difference data can be checked by comparing twice the difference in the calculated log likelihoods before and after iterative changes in the parameters. Because twice the difference in the log likelihood is approximately a Chi-square statistic, a Chi-square test can be used to determine whether a significant improvement has been made in reducing residual error in the model fit. Similarly, a Chi-square test can be used to establish the number of dimensions for model equation (2) by taking twice the difference in the log likelihoods of the model with k and with k-1 dimensions (k = 2, 3,... until Chi-squared is not significant at the 0.05 probability level). The number of degrees of freedom for this Chi-square test is n-k, where n equals the number of items judged.

So far, the difference observations of only one judge have been considered. It is also possible to calculate a group standard error estimate by summing the squares of the differences

$$\log d_{ijr} - \log d^{*}_{ij}$$

for all pairs of items ij and for all subjects r. Similarly, the log likelihood of the fit of model Equation (2) can be maximised for a group of subjects. It then becomes interesting to calculate the standard error estimate for each subject relative to the model, in order to calculate the standard error estimate for each subject relative to the model predicted differences d^{*}_{ij} for the group. A high individual standard error is an indication that the subject's responses are inconsistent with the assumption of a lognormal error distribution, or that the responses do not agree with the group model. Since inconsistency is a possible reason for eliminating a subject, it is desirable to consider these two sources of deviation from the group model separately. A quantile error plot, as shown in Figure 1, can be used for this purpose. For a subject r who is in agreement with the group model the normalised residuals

$$\frac{\log d_{ijr} - \log d^{*}_{ij}}{\sigma}$$

will have a normal distribution with mean zero and unit standard deviation. If this normal curve is cut into numbered quantiles of equal area and the normalised residuals are ordered according to size and plotted against their corresponding quantile numbers, they should form a line passing through zero with unit slope.

The quantile plot of an inconsistent subject will show a broad scatter of the normalised residuals about this line, while the normalised residuals of the subject who is consistent, but disagrees with the group model, will show a distinct trend away from this line. The quantile error plot also serves to weed out certain data recording errors which appear as spurious values (see Figure 1).

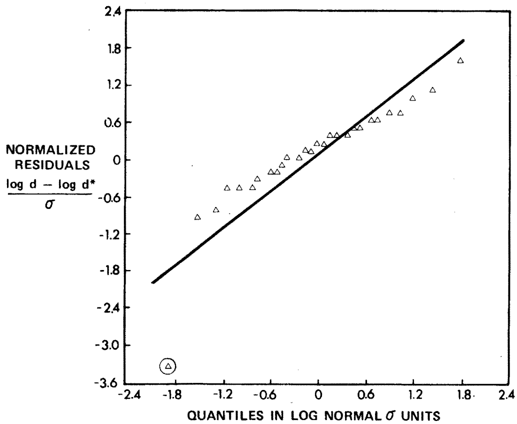


Fig 1—Quantile error plot of normalized residuals for a judge showing a mild deviation from the group difference model. The circled point is a deviation caused by a data recording error.

Testing procedure

Since most print quality judgements are made under ambient illumination conditions, it was decided to specify diffused fluorescent lighting, using common cool white tubes in the absence of daylight, for the subject evaluations. Print samples to be examined were mounted on neutral grey card stock and placed on a neutral grey examination table. Subjects were permitted to pick up the samples and examine them as they would reading materials. Each sample was marked with a number or letter. Judgements were recorded in a prepared booklet (see Figure 2) which instructs the judge to:

- indicate with a check mark on a continuous scale the degree of difference seen in each pair of sample prints.

- indicate with a check mark on a signed scale which print in each pair is preferred and to what degree.

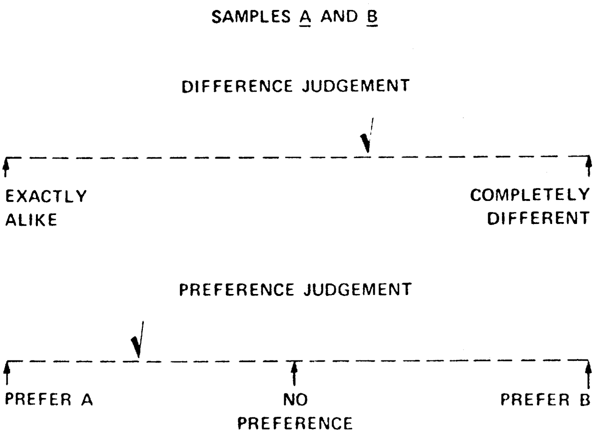


Fig 2—Stylized page from the prepared booklet used to record difference and preference judgements.

It should be noted that the judges are not instructed as to which factors should be used in making their judgements of difference and preference. MDS provides a means of extracting the statistically significant factors from the data, scaling these factors, and establishing their relative importance in the overall judgements of print quality.

The number of judgements should be kept to a minimum in order to avoid fatiguing the judges, but more degrees of freedom for statistical analysis are created when more samples are used. Eight sample prints have been found to give a reasonable compromise that allows the testing to commence with ten practice judgements involving all of the sample prints. The practice judgements permit the judges to stabilise their use of scales and familiarise themselves with the sample prints before starting on the judgements that are used for the model. Eight sample prints involve twenty eight paired comparisons, and since the difference and preference scales are divided into twenty five units, a great deal of detailed information is generated about the samples by each judge.

The difference and preference scales are not numbered in order to avoid possible numerical biases of the judges. The positions of the check are converted later to numerical values and entered into the Multiscale(11) computer program as a lower triangle matrix of differences. The preference information is treated similarly and entered, along with the difference information, into a hybrid computer program described in the section titled "Joint Analysis of Difference and Preference".

With the exception of the wire-mark study, all samples were printed in full commercial scale. Eight standard newsprints from mills in wiuely separated geographical locations in Canada and the United States were used in each of these trial printings, and in all trials the Paprican test print was used (Fig 3).

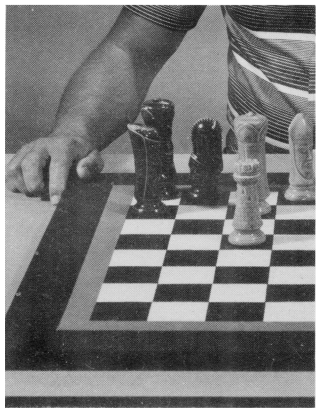


Fig 3—Paprican test print.

Printers were selected from the pressrooms involved and from pressrooms using the same printing process as used to create the sample prints. Specialists who judged print quality regularly were selected from the paper industry. Advertisers were generally recruited by the publishers and were thus familiar with judging print quality in the newspaper concerned. Readers were selected on the basis of their not having any specialised knowledge of paper-making or printing. All subjects were screened for colour blindness.

Wire-mark study

The forming fabrics used in paper-making leave indentations in the surface of the paper at the points where fabric weft and warp strands cross. The periodic pattern of these indentations is termed wire-mark and it can cause the same periodic pattern to appear in the printed image.

Eight forming fabrics used in the manufacture of newsprint were selected for a simple investigation of the effect of fabric weave on perceived wire-mark in solid letterpress prints. Paper was made on the eight fabrics in a British sheet machine using a groundwood/sulphite newsprint furnish. The handsheets were couched and pressed according to CPPA standards and dried with the top sides against the drying discs. After calendering to 80 Bendtsen roughness, the wire sides were printed with solid plates to 0.85 print density in a GFL letterpress tester.

The mesh and weaves of the fabrics and the percentage of ink transfered to the paper from the printing plate are listed in Table 1.

Sample	Weave	Mesh	% Ink transfer
1	5 shed (long warp up)	74x67	58.5
2	4 shed (long warp up)	84x56	58.2
3	4 shed (long warp up)	74x54	63.1
4	4 shed (long warp down)	74x50	60.2
5	5 shed (long warp up)	84x58	61.5
6	Duplex	108x65	65.8
7	Duplex	160x94	65.2
8	Bronze (broken twill)	68x53	65.0

Duplex refers to a two-layer fabric. A rough comparison with the mesh of a single-layer fabric is provided by dividing the number of warp strands by two.

Table 1
Wire-mark sample identification

The printed samples were mounted and judged according to the method described in the preceding section. Seven specialists in paper-making made the judgements. A Chi-square test of the log likelihoods for the one and two dimensional group MDS models (whose log likelihoods were 55.3 and 67.9, respectively, with 6 degrees of freedom) showed the latter to be significantly better at the 0.01 probability level. Comparison of the two and three dimensional group MDS models showed that the latter was not significantly better at the 0.05 probability level (log likelihoods 67.9 and 71.4, respectively, with 5 degrees of freedom). Thus the two dimensional model was selected. The group standard error estimate was 0.429 and twenty six parameters were used in fitting the individual difference data, leaving 170 degrees of freedom.

The individual standard error estimates, quantile error plots, and scalar and power correction parameters, V and p, were checked, and all seven judges were found to be highly consistent.

The final MDS configuration estimates for the eight printed samples are listed in Table 2.

Sample	co-ordinates x	y
1	38.06	-4.70
2	-6.87	-19.60
3	-20.75	-16.28
4	63.45	-12.12
5	-32.94	7.63
6	15.78	41.95
7	-40.40	-5.20
8	-16.74	8.31

Table 2
MDS final configuration estimates

When the print samples were laid out in this configuration and inspected by the specialists, it was apparent that the intensity of wiremark impression increased in a direction about -60 degrees from the x dimension axis and that the wire-mark pattern had a tendency to show more lines in a direction about +45 degrees from the x dimension axis. In order to confirm this observation a video scanner was employed to measure the root mean square (RMS) about the mean optical density, scanning in a direction perpendicular to the principal orientation of the wire-mark in the printed samples.

Multiple linear regression of the RMS values on the x and y co-ordinates of the samples resulted in a significant R^2 of 0.79 (significance at the 0.05 probability level for 5 df = 0.70). The direction cosine calculated from the regression coefficients corresponds to an angle of -57 degrees to the x dimension axis in the sample configuration. Regression of the percentage of ink transferred from the printing plate to the paper samples (listed in Table 1) onto the configuration co-ordinates bordered on significance at the 0.05 probability level. The direction cosine corresponded to a direction opposite to that of the RMS. This suggests that the non-transfer of ink to wiremark depressions left in the surface of the paper is the main cause of the differences in non-uniformity of the printed samples.

The degree to which lines appeared in the wire-mark pattern was measured with an image analyser. A 600 x 300 μm window was scanned over the prints with the window aligned with the principal direction of the wire-mark pattern. A modified Hough transform[12] was used to convert the measurements of the total area composed of lines (A_{L+N}), and the total area having no lines (A_N), to a single `linearity' value for each of the prints:

$$\text{linearity} = \frac{A_{L+N} - A_N}{0.5\ (S_{L+N} + S_N)} \tag{5}$$

where L+N refers to the areas having both lines and normal image noise, and N refers to areas having only image noise. S is the standard deviation of the mean value of the areas measured. The calculated `linearity' and RMS values are listed in Table 3.

Sample	RMS about mean OD	Linearity
1	0.168	0.932
2	0.173	0.507
3	0.165	0.335
4	0.176	0.878
5	0.158	0.066
6	0.159	1.268
7	0.158	0.466
8	0.152	0.40

Table 3
RMS and Linearity measurements on wire-mark prints

Regression of the linearity values onto the MDS co-ordinates resulted in an R^2 of 0.76 and a direction cosine corresponding to a direction of +48 degrees to the x dimension.

The wire-mark intensity and linearity axes were then applied to the MDS configuration. The configuration is shown in Fig.4 rotated so that the non-uniformity axis is horizontal (the model equation (2) is rotationally invariant). Ellipses around the mean configuration estimates are the 95% confidence regions for the locations of the eight prints(13).

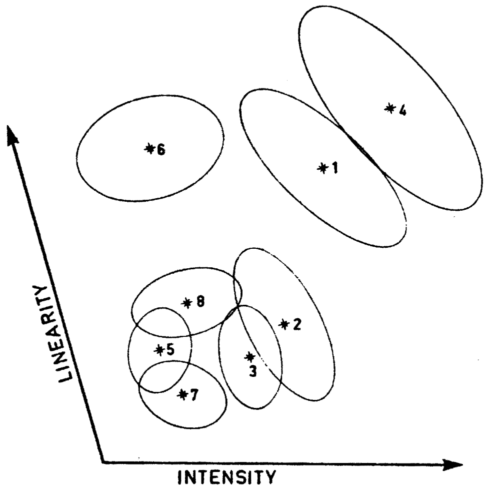


Fig 4—MDS configuration for papermakers' judgement of wiremark in eight sample prints printed on paper made on different forming fabrics showing the orientation of the identified subjective axes. Ellipses are 95% confidence regions for the location of the eight sample prints.

Preferences were also recorded and processed jointly with the difference information. This technique is discussed in the next section. All of the judges preferred a lower intensity and less linearity in the wire-mark pattern. It is well documented in the psychological literature that people are very sensitive to orientation in textures(3) and are much more disturbed by noise in an image if that noise has a textural orientation(14). Since linearity, which imposes orientation on the wire-mark patterns, is virtually independent of the wire-mark intensity (i.e. the linearity and intensity axes in Figure 4 are almost orthogonal) the perceived wire-mark can be lessened by simply avoiding fabrics which result in linear patterns. This can be accomplished without compromising the drainage characteristics of the fabrics in any way: this cannot be said for the other recognised way of reducing wire-marking, which is to use a finer wire.

Joint analysis of preference and difference

As mentioned previously and shown in Figure 2, judges are also asked to indicate on a signed scale (-12 to + 12) the degree to which they prefer item i to item j. This information can be reduced to a salience for each item i:

$$u_i = \frac{1}{n} \sum_{j=1}^{n} p_{ij} \tag{6}$$

where u_i = salience for item i
n = number of items
p_{ij} = preference for item i over item j
and p_{ij} = $-p_{ji}$
p_{ij} = 0 when i = j

A simple linear model for the preference of a judge for item i over item j can then be constructed from the calculated saliences:

$$p^*_{ij} = u_i - u_j \tag{7}$$

where p^*_{ij} is the model predicted value of preference for item i over item j.

In the cases discussed in this article there are eight prints to be judged, making a total of twenty eight pairs. Preference judgements for thirteen of these pairs are involved in the calculation of each p^*_{ij}. Thus a great deal more information is generated about the preferences of each judge than would be the case with a direct rating of salience for each item separately. The consistency of the preference judgements can be checked by comparing the model predicted preferences p^*_{ij} with the observed preferences p_{ij}. This can be expressed as the correlation between the p_{ij}'s and p^*_{ij}'s, which can in turn be used to calculate an F-ratio to test whether the relationship is significantly better than random. Thus, a high correlation (a highly significant F-ratio) indicates that the judge is consistent and that the model predicted preferences are statistically reliable.

The model predicted preferences can be linked to the co-ordinates of the items in their multidimensional configuration by a scalar product model.

$$p^*_{ij} = \sum_m a_m (X_{im} - X_{jm}) \tag{8}$$

where a_m = salience for dimension m.

The dimension saliences in the scalar product model thus define a preference direction in the MDS configuration.

Scalar and power corrections can be applied in Equation (8) as they were in Equation (2) to match the style of response of each judge when using the preference scale. The dimension saliences of different judges may then be compared. The dimension saliences may also be taken as dimensional components of a preference vector in the sample space. The length of this vector is an indication of the overall strength of the judge's preference and its direction is that in which his preference in the sample space increases indefinitely.

Since the above scalar product model of preference and the difference model Equation (2) share the same co-ordinates for the judged items the models can be fitted jointly. Ramsay[15] has described how the scalar product model of preference and the difference model can be jointly fitted using the maximisation of log likelihood approach. The preference judgements thereby add considerably to information about the configuration of the samples. The number of degrees of freedom (the total number of judgements less the total number of parameters used in fitting the model) in the joint model of preference and difference is about twice that for the difference model alone. This increases the confidence with which hypothesis tests can be performed. For example, the dimensionability of the joint model is decided using Chi-square tests on twice the difference in log likelihoods of the model in k and k-1 dimensions. The number of degrees of freedom for this Chi-square test is N=n-k, where n is the number of judged items (eight in the cases studied here) and N is the number of judges. This form of joint preference and difference analysis (PDA) has been used in the print quality evaluations described in the following sections.

Letterpress print quality study

A commercial printing trial was conducted on a stereo letterpress unit as part of the normal production of a daily newspaper. The Paprican test print was run as a black and white captioned ad. for subscriptions to the newspaper. The reverse side was printed with a grocery ad. having red and black solids. Eight rolls of newsprint from different North American suppliers were brought to the press unit running the test print. Only three thousand copies were printed between roll changes in order to avoid plate wear and lint build-up. Ink level on the printing plate was set during a warm-up run of the press and remained constant throughout the trial.

Copies bearing the test print were collected for each of the eight papers and mounted for judging. White paper samples were cut from the rolls after the run and sealed in bags for later laboratory print quality testing and physical analysis. The prints were judged by printers, technical specialists from the paper industry, advertisers, and readers.

The judgements of each group were entered into the program for joint analysis of preference and difference (PDA). Chi-square tests of the twice the difference in log likelihoods of the PDA models for different dimensionalities showed that a three dimensional model was appropriate for the groups of advertisers and printers, while a two dimensional model was appropriate for the groups of readers and paper-makers.

Visual examination of the PDA configuration of the sample prints for the printers' judgements suggests that mottle (perceived non-uniformity in constant tone areas of the prints), show-through of the grocery ad. from the reverse side of the prints, and contrast were factors. Therefore, corresponding image analysis measurements were made on the judged prints.

RMS non-uniformity measurements on the judged prints, and laboratory print tests for uniformity on the white paper samples, such as the Larocque number[16], IGT print density with 3 μm thick ink film[17], and NPIRI wedge point of continuous coverage[18], had R^2 values (0.77, 0.95, 0.86 and 0.89, respectively, 5 df) significant at the 0.05 probability level or better, and direction cosines locating the mottle axis in the same plane to within 10 degrees.

Image analysis measurements of show-through, and NPIRI wedge show-through[18] on the white paper samples had R^2 values (0.78 and 0.83, respectively, 5 df) significant at the 0.05 probability level or better, locating the show-through axis in the same plane to within 5 degrees. Measurements of print density in the dark tones of the judged prints located the contrast axis along one dimension axis (R^2=0.60, significance at 0.05 probability level = 0.50, 6df). This configuration was then rotated so that the mottle and show-through axes were in the xy plane and mottle and contrast axes were in the xz plane as shown in Figure 5 for the two planar views of the three dimensional sample space. The preference vectors for the fourteen printers are also shown.

It is apparent that the three axes are not independent. Sample prints having greater non-uniformity, or mottle, also have higher show-through and lower contrast. The relationship between contrast and mottle in the letterpress prints is in close agreement with that found by Parush[19] on a broader selection of letterpress prints.

While several of the sample prints cannot be distinguished by the printers (e.g. 4 and 6, and 1 and 8 in the mottle show through plane) number 5 is very distinct in having higher mottle and show-through and lower contrast than the others. Though the preference vectors vary, all printers want a combination of less mottle and show-through, and higher contrast. Thus, print sample 5 is the least preferred and print sample 1 the most.

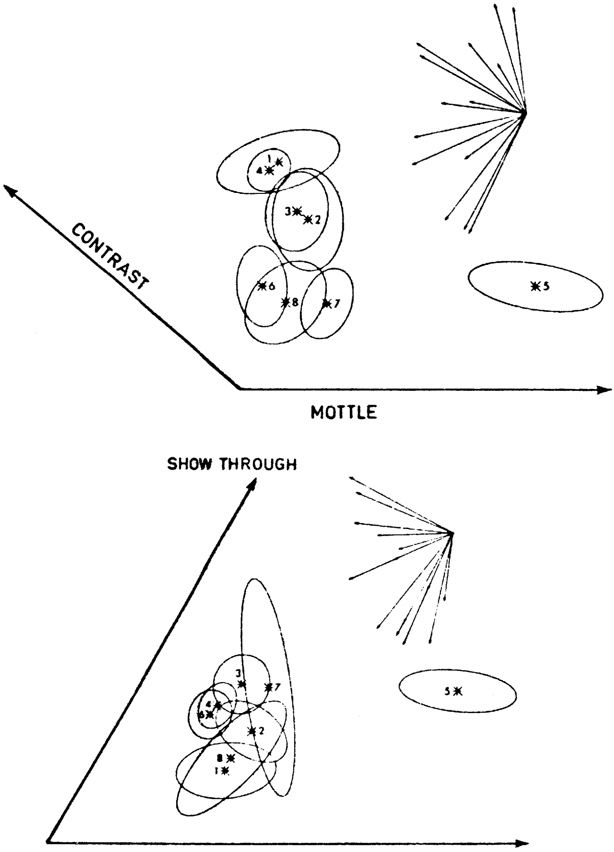


Fig 5—Three dimensional PDA configuration for printers' judgements of print quality in commercial letterpress prints printed on eight different newsprints. The xy plane is shown below and the xz plane above. Ellipses are 95% confidence regions for the locations of the eight prints and the vectors show the direction and strength of the preferences of each judge.

It should be noted that a three dimensional model may be required to describe the difference and preference judgements of a group of people who individually use only two factors to judge the prints. As can be seen in Figure 5 some of the preference vectors are virtually perpendicular to the mottle, show-through, and contrast axes. Thus, some printers are relatively insensitive to mottle while others are insensitive either to show-through, or to contrast. A listing of dimension saliences provided by the PDA program can be useful in determining which judges make use of all three dimensions, and which use combinations of two of the three.

Some caution should be exercised in generalising the saliences of a judge, or group of judges beyond the context of the evaluated print samples. For example, in a previous printing trial on the same printing press three factors were found to be significant in the judgement of the prints by the four groups of judges: mottle, contrast and paper colour. Unlike the trial described above where a grocery ad. containing red and black solids backed the test print, only text was printed on the reverse side of the test prints. Thus, the show-through was not sufficiently severe relative to the mottle and lack of contrast in the test prints to be a significant factor. A factor which is generally more subtle, paper colour, was used by judges in all four groups.

A simple way to appreciate the sensitivities of the four groups who judged the sample prints is to take an average of the preferences for each group and compute the components, or saliences, for each of the factors used in judging the prints. Figure 6 is a histogram of the absolute values of these saliences normalised so that their sum is unity for a combination of the above two stereo letterpress trials. Mottle was the major concern of all four groups while paper-makers were more concerned about the paper colour and printers were more concerned about contrast. This is not surprising since paper-makers control the paper colour, and printers control the contrast.

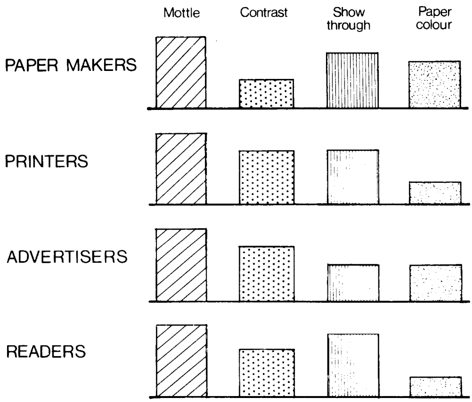


Fig 6—Histogram of the absolute values of the saliences of four groups who judged prints from two stereo letterpress printing trials normalized so that the sum of the saliences for the four subjective factors of significance equals unity.

In this and subsequent studies involving newsprint colour the b* (yellow-blue axis) of the CIE L*a*b* colour space[20] was found to have significant correlation with perceived paper colour differences in the same prints while saturation and dominant wavelength did not. Generally, preferences were for a bluer, (more correctly, less yellow) newsprint.

Thresholding and compounded subjective print quality factors

The subjective print quality factors which have been discussed thus far can be described by continuous scales. However, certain print quality defects only occur or are perceived after a certain threshold is surpassed. Linting is an example of this type of factor.

A group of technical specialists from the paper and publishing industries judged the Paprican test print printed on eight different newsprints in an experimental, light weight, polymer-plate letterpress unit. Multidimensional scaling of their difference and preference judgements showed that three factors had been used. The first axis was clearly mottle since the following measurements were closely aligned in the regression analysis:

	R^2
RMS on Judged Prints	0.98
Larocque Number[16]	0.89
GFL Ink Requirement[21]	0.85
Parker Print Surf Roughness	0.82
Typotest[22]	0.75

The third axis was significantly correlated with laboratory show-through tests. However, the second axis could not be identified by regression analysis.

Lint on the printing plate had been noted during the printing trial, necessitating washing the plate after each roll change. This lint appeared as white specks to a pronounced degree in sample prints 4 and 7, and to a lesser degree in the intervening prints 5 and 6 (the order of running was 1 to 8). IGT pick tests[22] showed that samples 1 and 3 had the highest surface strengths and samples 4 and 7 the lowest surface strengths (72, 80, 33 and 44 cm/sec, respectively). However, the correlation between surface strength and the co-ordinates of the samples along the lint axis shown in Figure 7 is not quite significant (R^2 = 0.46, 6 df.). Linting does not become severe until surface strength drops below a certain threshold. The orientation of the lint axis is therefore not precisely determined. However, as shown in Figure 7 the preference vectors point away from the samples which showed lint speck.

Difficulty in identifying and locating a subjective print quality axis also occurred in a polymer-plate letterpress printing trial at a daily newspaper. Multidimensional scaling of the printers' judgements of the Paprican test print printed on eight different newsprints showed that two factors had been used. The first axis was easily identified as mottle (R^2 = 0.83 with RMS nonuniformity). However, while inspection of the sample prints suggested that the second axis was show-through, the R^2 (0.55, 5 df) for the image analysis measurement of show-through was below significance at the 0.05 probability level.

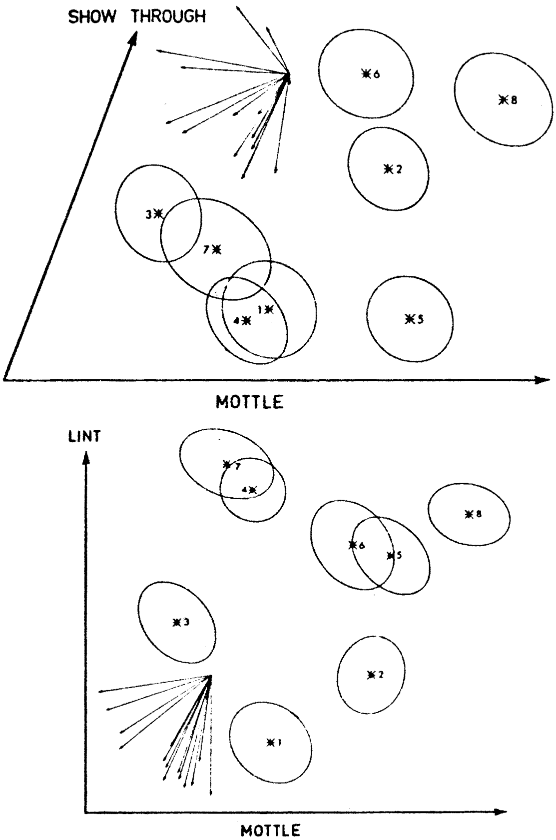


Fig 7—Three dimensional PDA configuration for technical specialists' judgements of print quality in prints printed on eight different newsprints in an experimental letterpress unit. The xy plane is shown below and the xz plane above. Ellipses are the 95% confidence regions for the locations of the eight prints and the vectors show the direction and strength of the preferences of each judge.

Closer examination showed that some of the prints alsc exhibited set-off. Set-off is a form of spurious imaging sometimes difficult to distinguish from show-through when the reverse side of the printed sample cannot be examined (as is the case for the mounted sample prints). Measurement of GFL set-off(21) on the corresponding white paper samples had a barely significant R^2 value (0.69, 5 df). The direction cosines from the regression analysis have been used to locate the show-through and set-off axes shown in Figure 8.

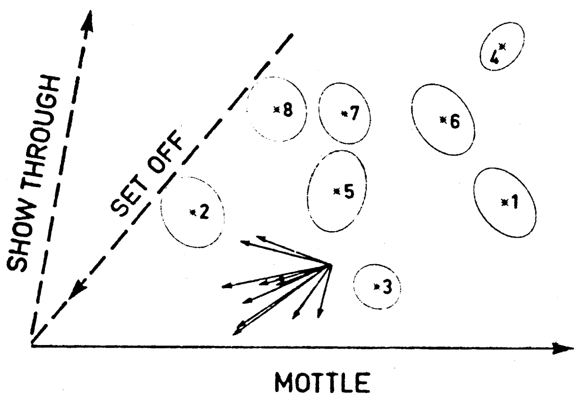


Fig 8—Two dimensional PDA configuration for printers' judgements of print quality in commercial polymer-plate letterpress prints printed on eight different newsprints. Ellipses are 95% confidence regions for the location of the eight prints and the vectors show the direction and strength of the preferences of each judge.

It is apparent that it was difficult to distinguish set-off from show-through in the prints, and since they have opposing trends in the sample space, it is not surprising that both show relatively low R^2 values.

Discussion

Any statistically-based model should be checked for intelligibility in order to guard against fortuitous correlations. For example, if the Chi-square test indicates that the highest dimension in the model is significant at the 0.05 probability level, but the subjects who judged the samples cannot identify the variables in that dimension, collapsing the model by one dimension should be considered. Similarly, more than one physical test should be used to identify each axis in the subjective configuration, and preferably, these physical tests should measure the relevant paper property in different ways. For example, image analysis measurement of RMS non-uniformity, roughness tests, the NPIRI wedge ink film thickness required for continuous coverage(18), and the typotest(22), all measure the non-uniformity of the paper surface using different principles. The direction cosines of these physical tests when regressed onto the configurational co-ordinates should agree closely if the model is to be accepted as consistent.

Eight sample papers is a practical number from the point of view of conducting a commercial printing press trial. Eight sample prints can also be judged for difference and preference without fatiguing the judges and without taking an excessive amount of time (typically about twenty minutes for an experienced judge and about forty minutes for an inexperienced judge). The eight sample prints result in twenty-eight judgements of difference and twenty-eight judgements of preference for each judge; in a PDA exercise this generates an adequate number of degrees of freedom for the determination of a group configuration model with a relatively small group of judges. However, five degrees of freedom for comparing the R^2 values for the regression of different physical tests onto the co-ordinates of the PDA configuration of the sample prints is a limitation of using only eight samples.

One way to increase the number of samples, n, without increasing the number of judgements, or necessarily changing the number of degrees of freedom for determining the PDA configuration, is to employ an experimental design for the nXn-1 matrix of difference and preference judgements. The model can then be used to estimate the missing judgements. However, if the number of samples that can be created is limited, as in the case of a commercial printing trial, pooled normalised values for a physical test and values for the corresponding subjective factor from different sets of judgements can be used to generate R^2 values with more degrees of freedom. For example, in two commercial printing trials Parker Print Surf, Bendtsen, and Sheffield roughness tests on white paper samples were correlated with corresponding subjective mottle values. The coefficients of determination (R^2) were:

	Trial 1	Trial 2	Pooled Trials
Parker Print Surf	0.820	0.915	0.875
Bendtsen	0.453	0.414	0.434
Sheffield	0.507	0.408	0.458

While it is possible to discriminate the R^2 values according to the level of significance it is not possible to reject the hypothesis that the R^2 values for the three roughness tests do not differ within each trial. However, by pooling the results of the two trials a sufficient number of degrees of freedom is obtained to show that the Parker Print Surf is significantly better related to perceived mottle than either Bendtsen or Sheffield at the 0.05 probability level.

Non-parametric tests such as the Spearman rank correlation can be used to test whether one physical test is significantly better than another in predicting a particular subjective factor if a sufficient number of trials is performed, or replicate evaluations are made. This can be done without increasing the number of test samples to be judged.

Summary and conclusions

Multidimensional scaling is a powerful tool for identifying and numerically scaling the factors used in subjective evaluation of paper and print quality. Joint preference and dissimilarity analysis of all sample pairs is a metric scaling technique which can accommodate different styles of response used by judges, and which generates a large number of degrees of freedom for hypothesis testing. Among other forms of output, this analytical technique provides a multidimensional configuration of the judged samples and a list of the dimension saliences of each judge. Physical measurements of related properties of the samples can be regressed onto the co-ordinates of the samples in the multidimensional space for purposes of establishing the identity of the subjective factors used to judge the samples and to determine the degree to which physical tests can be used to predict subjective judgements. The dimension saliences of each judge can be represented as a preference vector in the sample configuration whose length indicates the strength of the preferences of the judge. The preference vector is a guide to the most efficient manner in which to alter properties of a sample in order to make it more acceptable to the judge.

Multidimensional analysis of wire-mark in solid letterpress prints has shown that the degree to which lines appear in the wire-mark pattern is as disturbing as the overall intensity of the wire-mark pattern to technical specialists from the paper industry.

The joint analysis of preference and difference between commercial stereo letterpress prints has shown that paper-makers, printers, advertisers, and readers are chiefly concerned about mottle, or non-uniformity in tone. Indeed, mottle has been found to be the principal factor in all of the letterpress print quality evaluations examined here. This subjective print quality factor can be predicted with considerable accuracy by image analysis measurements of the RMS about the mean optical density in the prints and by laboratory printing tests such as the GFL ink requirement[21] and the Larocque number[16]. The Parker

Print Surf roughness tester also accurately predicts subjective mottle while the Bendtsen and Sheffield roughness testers do not.

Perceived newsprint colour was found to be significantly related to the yellow-blue axis value b* in the CIE L*a*b* colour space[20], and not significantly related to colour saturation.

Regression of IGT surface strength test values onto subjective co-ordinates for perceived lint in polymer-plate letterpress prints was not found to be a satisfactory means of identifying the subjective axis because of the threshold effect of surface strength on linting.

Perceived show-through in commercial prints can be predicted with laboratory print tests on white paper samples. However, some confusion of subjective axes may occur when both set-off and show-through are present in the prints to be judged.

REFERENCES

1. Ramsay, J.O., J. Opt. Soc. Am., 1968, **54**(1), 19-22.
2. Christian, J., `Psychological Differentiation and Definition of the Self. Multidimensional Scaling Approach', Ph.D. Thesis, McGill University, 1976
3. Julesez, B., in `Models for the Perception of Speech and Visual Form', (W. Walthem-Dunn ed.), 1964 MIT Press, Cambridge, Mass., 136-152.
4. Lyne, M.B., Tappi, 1979, **62**(11), 103-107
5. Torgerson, W.S, Psychometrika, 1952, **17**(4), 401-419
6. Torgerson, W.S, `Theory and Methods of Scaling', 1958, Wiley, New York
7. Shepard, R.M., Psychometrika, 1962, **27**(2), 125-140
8. Shepard, R.N., Psychometrika, 1962, **27**(3), 219-246
9. Ramsay, J.O., Psychometrika, 1977, **42**(1), 83-110
10. Ramsay, J.O., Psychometrika, 1977, **42**(2), 241-266
11. Ramsay, J.O.,`Multiscale. Four Programs for Multidimensional Scaling by the Method of Maximum Likelihood', 1978, National Educational Resources, Chicago.

12. Hough, P.V.C., `Method and Means of Recognizing Complex Patterns´, U.S. Patent 3069654, Dec. 1962
13. Ramsay, J.O., Psychometrika, 1978, **43**(2), 145-160
14. Cambell, F.W. and Kuikowski, J.J., J. Physiol., 1966, **187**, 437-445
15. Ramsay, J.O., Psychometrika, 1980, **45**(2), 149-165
16. Larocque, G., Axelrod, B. and Clark, S., Pulp and Paper Mag. Can., 1951, **52**(C), 166-173
17. TAPPI Useful method 505
18. Steinberg, S., Gen, C.F., and Hermann, W.K., Tappi, 1960, **43**(6), 539-552.
19. Parush, A., `The Relationships Among Perceived Contrast, Noise and Content in Printed Images´, M.A. Thesis, McGill University, 1980
20. Pauli, H., J. Opt. Soc. Am., 1976, **66**(8), 866-867
21. Scan Standards P35.72, and P36.77
22. Chiodi, R., Schoene, E., and Girard, P., ATIP, 1975, **29**(3), 89-95
23. TAPPI Suggested Method T499 su-64.

Acknowledgements

Appleton Wire is thanked for supplying the forming fabrics for the wire-mark study. The authors are grateful to the Miami Herald, St. Paul Pioneer Press, and the American Newsprint Publishers Association Research Institute for conducting the printing trials. The authors would like to thank the following staff members of the Pulp and Paper Research Institute of Canada. Emil Koller and Cory Malin for assistance in computing, Patrice Mangin and Julie Dalphond for performing the laboratory tests, Nam Nguyen for performing part of the image analysis measurements and suggesting the Hough transform approach to measuring wire-mark, and Derek Page for many useful discussions.

Transcription of Discussion

Discussion following paper 6.3, given by Dr. M. B. Lyne

Dr. J.R. Parker, Bowater Technical Services, UK

Bruce, that was most interesting. Information theory tells us that if you can detect a periodic signal then you can remove it by filtering, and it is interesting that the human eye is apparently unable to do this. More seriously, what happens if the relationship between `preference´ and `difference´ as seen by the observers is non-linear. In particular, what happens if `preference´, for example, shows a maximum. How would your analysis deal with this?

Dr. M.B. Lyne

To assess the effect of the response of the human visual system we investigated many different ways of quantifying the non-uniformity in print, using the image analyser. After using all sorts of complicated statistics we found that the RMS variation about the mean optical density gave the best agreement with perceived non-uniformity. Non-uniformity of spacing between samples does not matter, in so far as the programme can take care of it. If preference and difference do not quite coincide, there is no conflict, both contribute to the positioning of the samples. However, as is shown by the case of lint, linear regression cannot be used to identify a grossly non-linear factor.

Dr. M. Hussain, Abitibi-Price, Canada

Bruce, how do the variables you identified in your analysis relate to the parameters we were discussing earlier, formation and grammage distribution?

Dr. M. B. Lyne

The study that I have presented here was a very early one. The conclusion from the letterpress and gravure studies was that the most important factor in people´s assessment of print quality was non-uniformity. Dr. B. Jordan´s study that I mentioned earlier came to the interesting conclusion that it is the second order statistics of basis weight that have to be used to predict print uniformity.

Mr. O. Swanberg, Nordiskafilt, Sweden

Your identification of the effect of wire-mark is certainly very interesting, and agrees well with how we have found these fabrics to perform on newsprint. But I would disagree with you when you say that you cannot move towards finer wires because of reduced drainage. There is no theoretical reason why a finer wire should drain more slowly on a paper-machine, nor is there any experience to suggest it. There is indeed, a very strong trend throughout the industry to move to finer wires, primarily for better paper quality.

Dr. M. B. Lyne

I would have thought that smaller holes would mean slower drainage.

Mr. O. Swanberg

By finer wires, I mean wires with finer strands and more holes, so that the total open area is maintained the same.

Dr. M. B. Lyne

Ah, but when I said finer, I meant smaller holes.

Mr. J. A. Bristow, STFI, Sweden

How is your analysis affected if preference for the property under consideration, print density for instance, does not increase continuously, but passes through a maximum?

Dr. M. B. Lyne

Certainly, there are many properties of paper and other subjective values that reach maxima. Consider the analysis of taste for example, in which a tri-axial system, showing saltiness, sweetness and bitterness, would be common. Obviously, infinite amounts of any of these would be undesirable, and if you were to subject commercial chocolate bars to analysis you would find that they probably enclose a volume in this tri-axial space which includes the optimum values of each of the three parameters, the ideal point. I used throughout this analysis a so-called scalar product model which assumes the ideal point lies outside the sample space. I could have used an ideal-point model, such as that described in Ramsay's paper on preference-difference analysis, referenced in my preprint. Assuming, as I did, that the preference is for infinitely low mottle, infinitely low show-through, and high contrast, then the ideal point lies outside the cluster of actual print samples that I used. However, other paper properties, like colour, would be more suited to the ideal point model.