The Effect of Image Quality on the Identification of Cephalometric Landmarks

John S. McWilliam, D.D.S.
Ulf Welander, D.D.S., Ph.D.

The development of cephalometrics has led to a growing need for exact location of landmarks to improve quantitative studies of craniofacial growth, quantitative evaluation of treatment effects, and classification of cases. It seems logical to believe that cephalometric images with a high resolution and a high contrast should provide the best conditions for accurate landmark identification. Many other factors are, however, involved.

Björk[1] described errors arising from differences in projection between two films of the same individual. He also noted that the difficulties involved in locating different landmarks varied depending on the nature of the landmark examined. This was confirmed by Richardson[11] who noted that some landmarks were more reproducible vertically than horizontally and vice versa, e.g., subspinale and the anterior nasal spine.

Baumrind and Frantz[3] found that repeated identification of the same landmark on the same cephalometric image resulted in errors which were too great to be ignored. The magnitude of these errors in identification varied from landmark to landmark and the distribution of the errors for each landmark was usually noncircular and characteristic of that landmark. The factors influencing accurate identification were quoted as distinctness of structural detail, noise from adjacent structures due to superimposition of conflicting anatomical details, and conceptual judgment, a factor which is largely based on the past experience and radiological knowledge of the observer.

Midtgård, Björk and Linder-Aronson[9] found that an interval of one month between two registrations did not significantly affect the reproducibility of the landmarks examined; nor did the fact that duplicate registrations were made by two observers.

Baumrind and Frantz[4] studied the side effects of uncertain landmark identification and found these errors were significant when transmitted to angular and linear measurements. Since the magnitude of the landmark errors varied, it followed that all angular and linear measurements varied in reliability. In an earlier investigation Björk and Solow[2] reported that, since certain landmarks can be used several times in the course of an analysis, the greater the error of measurement, the greater the apparent correlation between the measurements. This spurious component may remain undetected even during tests of statistical significance. The influence of these errors on analyses based on superimposition has also been studied. It was found that they were sufficiently large to be of consequence in interpreting the changes taking place due to growth and/or treatment.[5]

It has been claimed by Rossmann and Wiley[12] that interpretation of radiographic images is dependent on radiological knowledge, pattern recognition, and physical image quality. According to this study, the physical image quality was the least important factor.

There is a direct correlation between the radiation energy absorbed by the patient during the exposure of a radiographic image and the physical image

From the Department of Orthodontics, Central Dental Clinic, Luleå, and the Section for Maxillo-Facial Roentgenology, Department of Diagnostic Radiology, Central Hospital, Boden, Sweden.

quality. A high degree of image resolution is, in other words, only achieved by increasing the patient dose. With this in mind the radiation dose should be limited to that necessary to achieve, not the best physical image quality, but an image quality which is adequate for accurate diagnostic purposes.

The recent use of rare earth phosphors in intensifying screens has made it possible to achieve a further reduction in the patient dose.[6] Until now the intensifying screens most frequently used have been based on $CaWO_4$, a fluorescent compound. Compared with medium intensifying screens of this type, the new rare earth phosphor screens make it possible to reduce the patient dose by a factor of about 4 at the same time maintaining the same image resolution and by a factor of about 6 with some loss in resolution.

The aim of this investigation was to test some of the new screens during cephalometric examination. It was hoped to establish whether images exposed with the most sensitive screens exceeded the tolerance limit regarding resolution, resulting in a deterioration in diagnostic image quality and reliability of landmark identification. Should this not be the case, high intensifying screens based on rare earth phosphors should be recommended for use in cephalometric examinations.

Material

The following screen-film combinations were tested: SE 2, SE 4, and SE 6 (CAWO, Philips) combined with Curix RP1 film (Agfa-Gevaert); Alpha 4 and Alpha 8 combined with XD film (3M). The SE screen series is equivalent to MR 200, MR 400, and MR 600 (Agfa-Gevaert) and TR 2, TR 4, and TR 6 (Kruppa, Siemens).[12]

Cephalometric images of a test object were exposed. The test phantom consisted of a dried human skull embedded in a plastic compound equivalent to soft tissues in attenuation of roentgen radiation. The phantom was mounted in a cephalostat and remained undisturbed until the entire series of images had been exposed.

The SE series was exposed using 60, 80, and 100 kvp, respectively. Since the Alpha screens produced a greater image contrast than the SE screens, this series was exposed at 80, 100, and 125 kvp to achieve approximately the same image contrast in both series. The mAs product was adjusted for each exposure to give an average film density of about 1 for the entire material. This was checked by scanning each film with a densitometer.

Film processing was standardized using an automatic developing machine.

Method

The different screen-film-kvp combinations were gathered into five identical series. The films (15) were arranged in ascending order of image quality according to their physical data. Five experienced orthodontists were asked to take part in the investigation as unbiased observers.

Landmarks

The landmarks to be identified were chosen with a view to studying as wide a variety in character as possible including both skeletal and soft tissue landmarks (Fig. 1).

The landmarks were clearly defined for each observer to eliminate any differences in opinion. The observers were then asked to identify each landmark by making a pin prick directly on the film. Each series was examined in numerical order and under standard conditions with a minimum of four hours between each film.

Evaluation

Using a positive image printed from the original series and with a transpar-

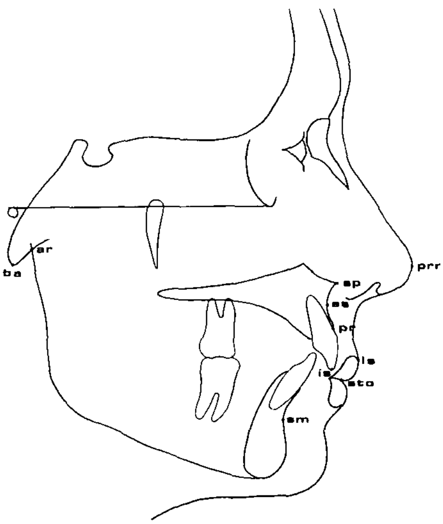

Fig. 1 Articulare (ar); basion (ba); incision superius (is); labrale superius (ls); prosthion (pr); pronasale (prn); spinal point (sp); supramentale (sm); subspinale (ss); stomion (sto).

ent millimeter square graphic inset covering the area to be examined, it was possible by a process of subtraction to orientate an identical system of coordinates over each image (Fig. 2). Each pin prick was then identified and given an x and y coordinate read to the nearest 0.1 mm using a calibrated magnifying glass with a magnification factor of 8. The collection of data was performed by one of the authors.

The coordinate system employed for registration was randomly selected. Due to the characteristic distribution of the estimations for each landmark, i.e., the varying pattern and orientation (Fig. 3a), the distribution along the x and y axes of the coordinate system is not representative of image quality. In a noncircular distribution the distribution along the long axis is more influenced

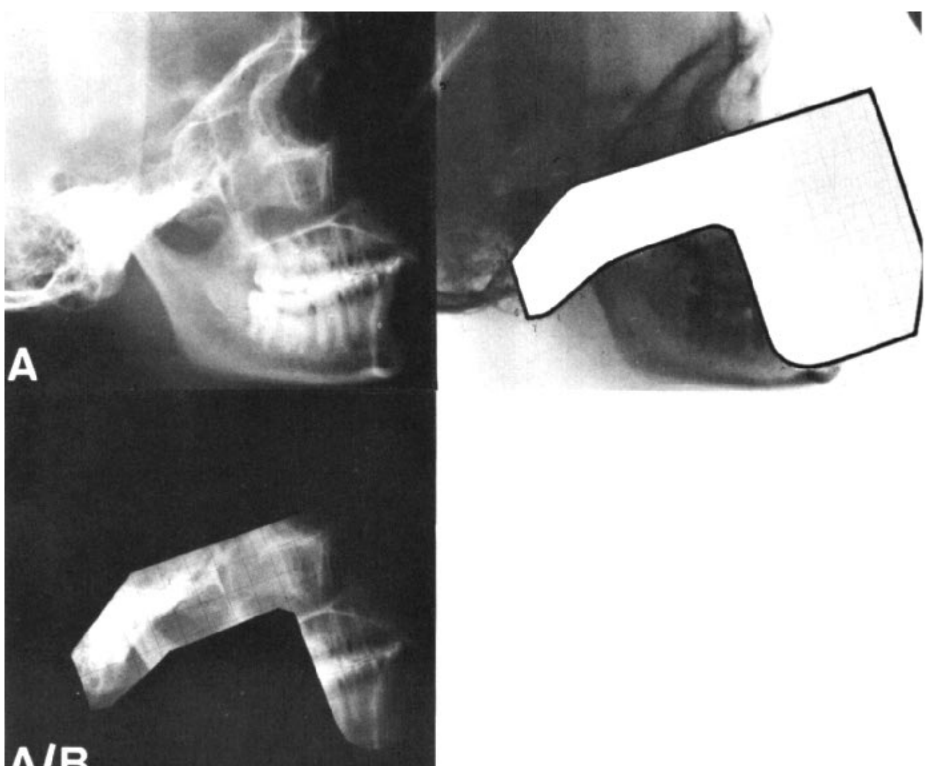

Fig. 2 Illustration of the cephalometric image (A) and the positive print template including a coordinate inset (B). Perfect superimposition (A/B) results in a total blackout of all structural details allowing a reproducible location of the coordinate inset.

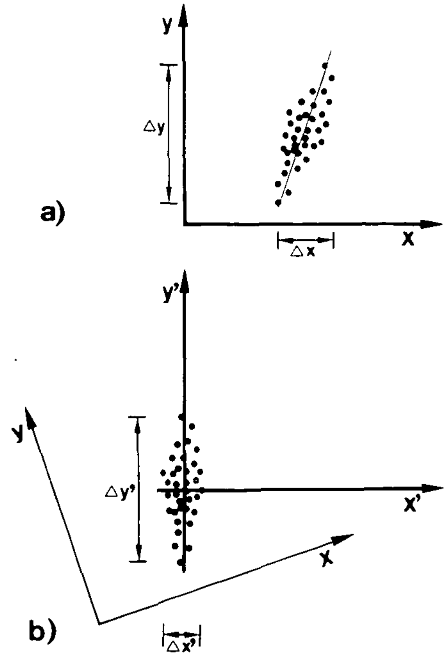

Fig. 3 Rotation and translation of a noncircular distribution of collected estimations for one landmark. The regression line becomes vertical and the mean lies over the origin of the new coordinate system. Note the change in the range, Δx and Δy as compared with Δx' and Δy', respectively.

by landmark identification than the distribution along the short axis. The latter should be the most representative measure of diagnostic image quality. To eliminate the error caused by noncircular distributions with random inclination to the coordinate system a regression analysis of the collected data for each landmark was performed and the inclination of the regression line was determined. By simultaneous coordinate rotation and translation all data could be uprighted around the origin of one and the same new coordinate system (Fig. 3b). In addition, the practice of translating the mean to the origin succeeded in eliminating systematic errors occurring between the observers. This mathematical adjustment of the data was computed using the following expressions:

TABLE I
Classification of coordinates into three groups according to their standard deviations.

Group	Standard deviations	Landmarks
I	0.19-0.21	ar x ' and y' coordinates
		is x' and y' coordinates
		sm x' coordinate
II	0.33-0.50	ba x' coordinate
		ls x' and y' coordinates
		pr x' and y' coordinates
		sm y' coordinate
		ss x' coordinate
		sto x' coordinate
III	0.57-0.96	ba y' coordinate
		prn x' and y' coordinates
		sp x' and y' coordinates
		ss y' coordinate
		sto y' coordinate

$$x' = (x - x_0)\cos\alpha + (y - y_0)\sin\alpha$$
$$y' = -(x - x_0)\sin\alpha + (y - y_0)\cos\alpha$$

where α denotes the inclination of the regression line, and (x_0,y_0) defines the origin of the new coordinate system.

Rogue observations or "outliers" were eliminated using a test of internal consistency.[10] Standard deviations were then calculated for each landmark within the new coordinate system. The x' and y' coordinates were treated individually and classified into three groups according to the size of each respective standard deviation (Table I).

Statistical analysis

Provided the x' and y' coordinates of each landmark are normally distributed, the quotient of the variance for the reference material and the various screen-film-kvp combinations should follow an F-distribution. The statistical analysis was carried out on this assumption.

The images exposed with the SE 2 screens were chosen as the reference images in assessing the other screen-film combinations since they, along with the Alpha 4 screens, were known to have the best physical properties. This comparison was performed for each group of landmarks.

Method Error

The error of the method employed in collecting the data was tested. Twenty double determinations were made of both x and y coordinates, the positive print template being readjusted between each registration. The variance of the method error was then analysed in relation to the variance of each group of coordinates as a whole to assess its importance.[8] It was found that the error variance was less than 8 percent of the total variance for the group of coordinates with the smallest standard deviation, Group I. This would indicate that the error of the method may have had little influence on the results obtained from this group. Regarding the other two groups, the error variance was found to be less than 3 percent of the total variance in each group. Consequently, the error of the method was of little importance in assessing these results.

Results

A statistical analysis based on the data collected from all five observers was confusing. Significant and insignificant results showed no correlations with expected results taking into account the known physical properties of the different screen-film combinations. It was noted, however, that one of the observers had a large number of rogue observations or "outliers." A statistical comparison between the observers revealed that the consistency in their observations varied. Of the five observers one showed a remarkably high consistency irrespective of the physical image quality, and one had results that were significantly inconsistent compared with the remaining four. A repeated statistical analysis based on the four observers showing consistent observations gave the following results.

From Table II it can be seen that Group I shows a significant difference between the identification of landmarks in the reference images and images exposed with Alpha 8 screens. This difference could not be registered at 125 kvp but was nevertheless strongly significant on examination of the group as a whole. Regarding images exposed with SE 6 screens, the difference, though apparent, was not as strong and individual analyses of images exposed at 80 kvp and 60 kvp showed no significant difference. The images exposed with Alpha 4 screens were not found to differ from the reference images neither as a whole nor at the individual kvp employed. This also applied to images exposed with SE 4 screens.

For Group II no significant differences could be noted between the reference images and those exposed with SE 6, SE 4, Alpha 8, and Alpha 4 screens.

TABLE II

A comparison between the reference material (SE 2) and the other screen-film combinations within each group. Differences are given for low, medium, and high contrast images.

Screen	Low contrast image†	Medium contrast image†	High contrast image†	Total
Group I				
SE 6	*	NS	NS	**
Alpha 8	NS	**	***	***
SE 4	NS	NS	NS	NS
Alpha 4	NS	NS	NS	NS
Group II				
SE 6	NS	NS	NS	NS
Alpha 8	NS	NS	NS	NS
SE 4	NS	NS	NS	NS
Alpha 4	NS	NS	NS	NS
Group III				
SE 6	NS	NS	NS	NS
Alpha 8	NS	NS	**	NS
SE 4	NS	NS	NS	NS
Alpha 4	NS	NS	NS	NS

†SE series exposed at 60, 80, and 100 kvp; Alpha series exposed at 80, 100, and 125 kvp.

The analysis of Group III gave similar results to those found in Group II. No significant differences in landmark identification could be found on comparing the reference images with images exposed with SE 6, SE 4, Alpha 8, and Alpha 4 screens with one exception. Those images exposed at 80 kvp using Alpha 8 screens showed a significant difference at the 1 percent level; however, this did not make itself manifest on examination of the group as a whole.

Discussion

In radiological diagnosis there are very few situations where it is possible to establish an optimum imaging system for both the necessary diagnostic information contents and the dose given to the patient. Cephalometric examinations represent one such situation since the diagnostic information gained from the examinations is dependent on the reliability of landmark identification. The present investigation has proved successful in objectively measuring the reliability of landmark identification. It has also shown that it is possible to establish a limit beyond which a further decrease in image resolution significantly increases the uncertainty in landmark identification.

To limit the investigation to an objective study of image quality as related to identification of certain landmarks it was necessary to minimize all other possible sources of error. A series of cephalometric images exposed without disturbing the test object eliminated errors due to differences in projection. Scanning of each film with a densitometer made it possible to adjust the exposure to achieve uniform density throughout the entire series. By placing each series in an ascending order of physical image quality it was hoped that the observers would not gain any positive assistance from previously examined images. The landmarks were discussed in detail with each observer to avoid any differences in opinion about definition. In addition, the practice of translating the coordinates so that the mean values became centralized over the origin of one and the same coordinate system was applied to eliminate systematic differences between the observers. This method is, to our knowledge, unique in connection with evaluation of cephalometric measurements. This can also be said of the subsequent rotation to a vertical position of the regression lines calculated for each landmark to study the distribution of the collected data to give the most reliable measure of diagnostic image quality.

Developing an accurate method of registering the identified landmarks was a difficult problem. This had to be performed so the expected minor differences in landmark identification between images exposed with different screen-film combinations were not obscured by the method of data collecting. Registration of the landmarks identified using the method of superimposition and subtraction was found to have an insignificant method error.

By separately analysing those coordinates having the lowest standard deviations, Group I, it was hoped that differences in diagnostic image quality would be most easily detected. This was found to be the case and a gradual deterioration in image quality could be reported with no significant difference for the Alpha 4 and SE 4 screens compared with the reference screens, a significant difference for the SE 6 screens (1% level), and a strongly significant difference for the Alpha 8 screens (0.1% level). These findings are in good agreement with the known physical properties of the screens. Image resolution at 80 kvp and 30 percent contrast has previously been established for the different screens: Alpha 4, 3.5

line pairs per mm, SE 2, 2.6 lp/mm, SE 4, 2.5 lp/mm, Alpha 8, 2.2 lp/mm, and SE 6, 1.8 lp/mm.[13] With regard to this parameter of physical image quality, images exposed with Alpha 8 screens should be superior to those exposed with SE 6 screens. The results obtained from this investigation suggested the opposite. There may be several explanations for this finding, the most probable being quantum mottle influencing the former images more than the latter thus reducing the diagnostic image quality as compared with the physical image quality. Furthermore, due to differences in sensitivity between the various screen-film combinations at different kilovolts, the contrast in the images is such that the resolution is not directly comparable at one single kilovolt value. The method error may also have influenced the results obtained. The gradual deterioration in diagnostic image quality seen in Group I could not be demonstrated in Groups II and III due to the masking effect of other more important factors than the physical image quality, i.e., the larger standard deviations which express greater uncertainty in landmark identification.

The interobserver differences which led to the elimination of one observer from the final statistical analysis illustrate the fact that such variations may often be greater than the effects of physical image quality. Previous investigations have examined the effects of other factors such as radiological knowledge, differences in projection, distinctness of structural detail, and noise from adjacent structures on the reliability of landmark identification. It would appear from this study that physical image quality is the least important of these factors. Landmark identification may be regarded as an application of pattern recognition. The results of this study can, therefore, be said to confirm the observation of Rossmann and Wiley[12] that physical image quality is less important than pattern recognition.

The effects of physical image quality on landmark identification became significant for those landmarks showing the narrowest distribution. It is questionable, however, whether this finding is clinically meaningful since other landmarks with a wider distribution are likely to mask these few specific and distinct points during measurements of angles and distances in the images.

Regarding the choice of kilovolts it may be said that those images exposed with high kilovolts (low contrast images) tended to be preferable. The soft tissues became more easily identified, a factor of increasing significance in contemporary cephalometric analysis. Furthermore, by increasing the kilovolts the patient dose is usually reduced.

The conclusions to be drawn are that high intensifying screens are to be recommended for use in cephalometric analyses. They help to reduce the patient dose during the exposure of images, in some cases by a factor as great as 6, and the sacrifice in image quality does not seem to jeopardize ordinary clinical demands on the reliability of cephalometric measurements.

Summary

The effect of rare-earth phosphor intensifying screens on diagnostic image quality was examined. Five observers were asked to identify ten landmarks on each of fifteen cephalometric images having varying screen-film-kvp combinations. The most sensitive screens were found to significantly affect diagnostic image quality and thereby the reliability of landmark identification. This finding was, however, only established for those landmarks having the lowest standard deviations and the

clinical significance is thus questionable. There would appear to be other more important factors than physical image quality involved in the reliability of landmark identification. High intensifying screens are, therefore, recommended for use in cephalometric analyses since they help to reduce patient dose and do not appear to sacrifice diagnostic image quality to the extent that the reliability of cephalometric measurements is significantly affected.

Dental School
Diagnosvagen
141 44 Huddinge
Sweden

REFERENCES

1. Björk, A.: The face in profile. *Swed. Dent. J. Suppl.* 40, 1947.
2. Björk, A. and Solow, B.: Measurements on radiographs. *J. Dent. Res.*, 41:672, 1962.
3. Baumrind, S. and Frantz, R. C.: The reliability of head film measurements. 1. Landmark identification. *Amer. J. Orthodont.*, 60:111, 1971.
4. ———: The reliability of head film measurements. 2. Conventional angular and linear measures. *Amer. J. Orthodont.*, 60:505, 1971.
5. Baumrind, S., Miller, D. and Molthen, R.: The reliability of head film measurements. 3. Tracing superimposition. *Amer. J. Orthodont.*, 70: 617, 1976.
6. Buchanan, R. A., Finkelstein, S. I. and Wickersheim, K. A.: X-ray exposure reduction using rare-earth oxysulphide intensifying screens. *Radiol.*, 105:185, 1972.
7. Cleall, J. F. and Chebib, F. S.: Coordinate analysis applied to orthodontic studies. *Angle Orthodont.*, 41: 214, 1971.
8. Linder-Aronson, S.: Adenoids: their effect on the mode of breathing and nasal airflow and their relationship to characteristics of the facial skeleton and the dentition. *Acta Oto-Lar.*, Suppl. 265, 1970.
9. Midtgård, J., Björk, G. and Linder-Aronson, S.: Reproducibility of cephalometric landmarks and errors of measurements of cephalometric cranial distances. *Angle Orthodont.*, 44: 56, 1974.
10. Pearson, E. S. and Hartley, H. O.: *Biometrika tables for statisticians.* Volume I. Cambridge University Press, 1970.
11. Richardson, A.: An investigation into the reproducibility of some points, planes and lines used in cephalometric analysis. *Amer. J. Orthodont.*, 52:637, 1966.
12. Rossmann, K. and Wiley, B. E.: The central problem in the study of radiographic image quality. *Radiol.*, 96:113, 1970.
13. SPRI: Röntgenkassettens förstärkningsskärmar; egenskaper, användning mm. An official Swedish SPRI publication, Project 6030. In press.