DAMAGED TEXT IN ORIGINAL

Text cut off in original

# THE ERGONOMICS OF VIDEO DISPLAY TERMINAL WORKPLACES IN INTERNATIONAL TELEPHONE EXCHANGES

By

T. G. MOORE

VOLUME 1 (of 2)

A DOCTORAL THESIS

Submitted in partial fulfilment
of the requirements for the award of
Doctor of Philosophy of the
Loughborough University of Technology.

FEBRUARY, 1981.

Supervisor: Professor N. S. Kirk
Department of Human Sciences.

VOLUME 1 (of 2)

NOTE Volume 1 contains the text of the thesis, Volume 2 contains the Addenda, consisting of References, Figures, Tables and Appendices.
This format allows easy reference to these Addenda while reading the main text of the thesis.

## SUMMARY

Since 1976 the author has worked with L.M. Ericsson Ab of Stockholm, Sweden, providing ergonomics advice relating to the design of operator positions for stored program controlled telephone exchanges manufactured by this company. The research was performed in three phases: a literature review, preliminary experiments to evaluate alternative equipment designs and finally a series of trials of the prototype under simulated call conditions.

The basis of the new operator position is the provision of a computer terminal, consisting of a datascreen or video display unit (vdu) and keyboard at each operator desk. These terminals enable the operator to enter call details into memory stores from which they can be 'read' by the computer, displayed at any operator position or routed to the charging system. By eliminating many repetitious and error prone number entries, by abolishing paper handling tasks completely and by automating certain call handling operations a more efficient and effective service can be provided to subscribers.

The literature review report, produced on completion of Phase I, summarised the ergonomics research relating to datascreens, keyboards, dialogues, workstations and environmental factors. As a result of the review the author was able to produce detailed ergonomics recommendations for many parts of the system.

Phase II consisted of a programme of experiments to provide the data required to resolve design conflicts remaining at the conclusion of Phase I. In successive stages a panel of experts reduced the potential number of datascreen designs to practicable numbers for controlled experiments. These datascreens, filters and screen treatments were systematically compared in experiments and a suitable design was selected.

In Phase III an experiment was conducted to evaluate the entire operator position under simulated call handling conditions. In addition to collecting objective measures of call handling speed and accuracy, the experiments also enabled visual fatigue to be assessed and call handling equipment and procedures to be judged by experienced operators under reasonably realistic conditions.

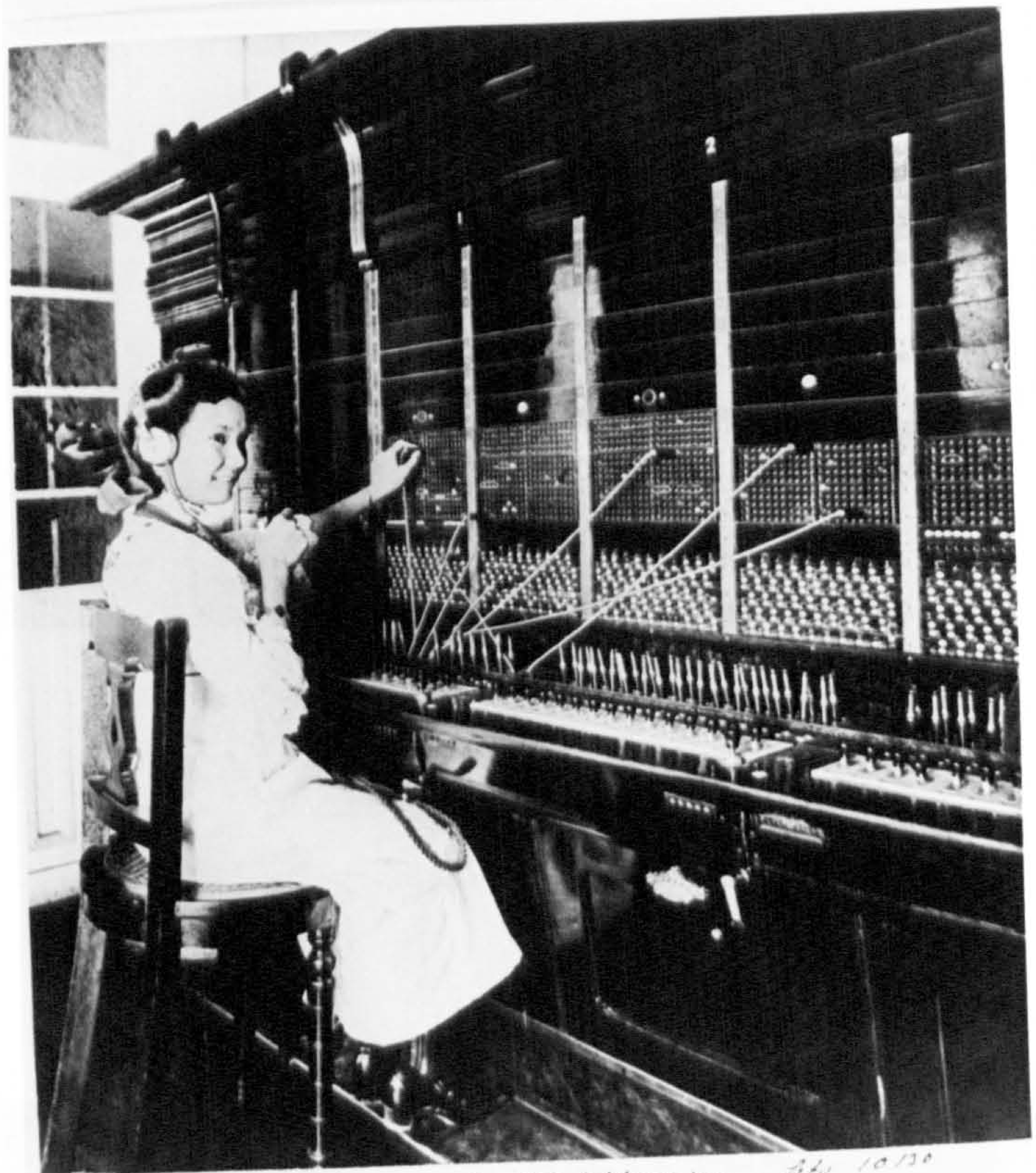
# ACKNOWLEDGEMENTS

The author would like to take this opportunity to acknowledge the assistance received from many sources during the period over which this thesis was prepared.

The research was supervised by Professor N. S. Kirk of the Department of Human Sciences in the Loughborough University of Technology. The author is especially grateful to him for his generous advice, encouragement and patience.

The research was based upon a project undertaken for L. M. Ericsson Ab., an international telephone equipment manufacturer with its headquarters in Stockholm, Sweden. The author would like to thank all members of staff at L. M. Ericsson who helped or took part in the project. In particular he would like to thank the following: Mr. L-E. Morell, Mr. B. Troili, Mr. W. Pegert, Mrs. E. Sager and Mr. L. Nielsson.

Of course, finally it is essential to thank those who exhibited great patience and understanding in the typing and preparation of this thesis. Particular thanks are given to Mrs. N. Johnson, Mrs. C. Dawson, Mrs. P. Stenhouse, Mrs. M. Symes and Mrs. J. Brewin.



Telefonstation i Hongkong, första utbyggnaden 1905. Induktorsystem, multipelbord med lampor för påringningssignaler och dubbel slutsignal, parallelljackar och parsnören.

Foto 10130

PAST

Frontispiece A. A cord keyboard in the 1900's.

PRESENT

FRONTISPIECE B. A cordless keyboard in the 1970's

FUTURE

FRONTISPIECE C. The ANE 403 switchboard for the 1980.

(Subject of this thesis)

CONTENTS LIST

CHAPTER C - PHASE II EXPERIMENTS TO SELECT DATASCREENS AND TREATMENTS

Page



## CHAPTER D - PHASE III PROTOTYPE TESTING USING SIMULATED CALLS

# CHAPTER A - GENERAL INTRODUCTION

## A.1 BACKGROUND

This thesis is based upon a programme of research instigated and funded by L.M. Ericsson AB (L.M.E.), an international company which manufactures and installs telephone and telecommunications equipment. The author and his supervisor worked closely with the company engineers based at the headquarters in Stockholm, Sweden.

While designing new Stored Programme Controlled (SPC) automatic switching equipment for international exchanges the engineering designers recognised the continuing need for human operators to handle certain types of telephone call (eg. person-to-person). Moreover, the engineers also appreciated that novel forms of operator console, incorporating visual display units (vdu's), would interface easily with these computer controlled exchanges. Aware of reports from Sweden and elsewhere, which described ergonomics difficulties with vdu based systems (eg. Östberg, 1974), and conscious of their own inexperience in designing screen based terminals, the engineers from L.M.E. approached Prof. Kirk and the author to provide ergonomics advice. Initially this advice took the form of a literature review and some preliminary experiments to establish an ergonomics specification. While the first system was being designed a simulator was built, which was initially used by the author to evaluate the design, and which was subsequently developed into an operator training aid. The literature review, preliminary experiments and simulated system evaluation form the basis for this thesis.

The operator system has now been sold to a number of telecommunications administrations around the world and the first system was completed in 1980.

## A.2 THE EVOLUTION OF TELEPHONE EXCHANGES

The telephone was invented in the early 1870's and its patent was taken out in 1876 by Alexander Graham Bell. The device was developed rapidly, and by 1920 there were 13 million telephones in the U.S.A. alone.

The first telephone handsets were simply connected in pairs but it is easy to demonstrate that direct connections between a number of handsets becomes unreasonable even for small numbers of sets. (e.g. Ten telephone sets require 45 connections if each is directly connected to each other). Furthermore, with such a system, each wire connection is used infrequently and cannot be considered cost effective. Telephone exchanges are necessary to act as concentrators and to conduct the switching operations which enable each instrument to be connected to any other instrument only for the time necessary to make the call. Furthermore, the exchanges contain equipment which enables single wires between such exchanges to carry a number of calls simultaneously.

The earliest exchanges were entirely manual. The subscriber was answered by an operator who routed the call via a patchboard. Such a system was very labour intensive, but it was a problem of privacy which lead a funeral director named Strowger to design the first automatic exchange switches. Strowger was losing business to a competitor whose wife, a telephone operator, was routing calls intended for Strowger through to her husband. The switch which Strowger designed enabled subscribers to dial directly to others without going through the operator and it proved so useful and reliable that its use spread rapidly, and many old exchanges throughout the world still use switchgear based upon the original design.

However, the Strowger switch was expensive and, since each switch contained its own control gear, this was also engaged during the time the lines were occupied. The Crossbar exchange designed at L.M. Ericsson in 1925 overcame these difficulties by separating the control functions from the switching functions. The Crossbar exchange contained a matrix of silver plated needles moved by solenoids and held in place by magnetic force. The next two generations of exchange operated on the relay principle. The first employed large relays and the next used reed relays which were smaller, required less operating force, and could be environmentally sealed. Alongside these developments in the design of the electro-magnetic/mechanical exchanges were advances in handling greater numbers of calls on each wire and the development of networks providing alternative routes to cater for peak loads. However, the cost of electromagnetic forms of switching continued to increase, whereas the prices of electronic forms of switches, and particularly electronic logic and control devices, fell dramatically. Since these new devices also offered greater flexibility and adaptability, stored program control (SPC) became the next logical step. This thesis describes the ergonomic aspects of the design of an operator terminal to interface with an SPC exchange.

New forms of exchange are now being designed incorporating electronic switching as well as electronic control. The system for which the operator position detailed in this thesis was designed was code named ANE and the code name for the console was ANE 403. Later it was found that many of the design features of this system could be incorporated into operator positions for these totally electronic exchanges.

## A.3 CHANGING FORMS OF OPERATOR CONSOLE

Paralleling the developments of new forms of exchange equipment (see previous section, A.2) there have been changes in the operator's role and in the form of console used for traffic handling. In the earliest exchanges the operator's prime function was to conduct all the switching operations. As switching operations were being increasingly automated, however, more operator assisted facilities were being introduced (e.g. credit cards). In consequence, the demand for operator assistance did not fall as dramatically as was expected. The present functions which are now handled by operators are dealt with in the next section (A.4) but this section (A.3) deals with the changes which have occurred in the form of console provided for operators.

### A.3.1 Patchboards

Even today many people think of the operator sitting before patchboards since these are still common in small company exchanges (see Frontispiece A). Most public exchanges are now very different (see A.3.2).

In a patchboard exchange the operator sat close to a large panel of sockets under which was mounted a horizontal surface covered in holes through which passed wires terminating in jackplugs which prevented the wires being pulled down through the holes. Below the horizontal surface the wires were looped around weighted pulleys which maintained a downward pull on the wires, returning the jackplugs to their resting place when not plugged into a socket. To make a connection the operator selected the appropriate jackplug and pulled it upwards before placing it into the appropriate socket in the switchboard. In addition, the operator was able to control the state of the connection, talking to either subscriber or listening to the call. Labels on jackplugs and sockets, together with selected auditory connections, and sometimes lights were provided to supplement the operator's memory when handling calls. In such an exchange the operator could supervise a number of calls at any one time, but these calls were directly wired through the operator position from inception to termination. All details pertaining to the call, including subscribers numbers, type of call, duration etc., were hand-written onto paper 'tickets! These tickets were used by the operator to store these details for later attempts to make a call which initially proved unsuccessful. Once the call had been made the ticket was then passed to clerks, who, in the earliest system, had to transcribe them and produce accounts by manual methods.

From a workstation viewpoint these patchboard consoles had a number of drawbacks. The high and wide panels of sockets required awkward and extensive arm movements to lift and insert the plugs leading to postural fatigue. Moreover, the size and location of these panels restricted social contact between operators and made it difficult for the operators to see common information displays. The presence of the weighted pulleys below the horizontal desk also made it difficult for operator's knees to be accommodated comfortably.

A.3.2 <u>'Cordless' exchanges</u>

With the advent of automatic exchanges it was no longer essential for operators physically to make the switching connections. The operator could now make these connections via an ordinary telephone dial. These so called 'cordless' exchanges are still the most common form of equipment used for manual assistance and traffic handling. (see Frontispiece B).

The operator consoles of a cordless exchange contain a telephone dial or numerical keypad, a number of switches to control internal switching operations (eg. speak to calling subscriber), a number of lights to indicate the state of lines and, in more recent years, digital displays to indicate the dialled numbers, metering units, etc. (See Frontispiece B). Calls handled through an operator position are associated with that position throughout their duration, and are supervised by the operator with the assistance of some automatic facilities such as time metering. In international exchanges the consoles have to be 'hard-wired' to take pre-determined types of call (e.g. those requiring a French speaking operator) and consequently operators are often required to move to another console when moving from one language group to another.

The cordless exchanges described in the above paragraph still require operators to complete paper tickets with call details, and subsequently to use these tickets for further call attempts, re-dialling the telephone numbers on each attempt. Computerised accounting systems have been introduced widely but these require keypunch operators to enter the call details manually into the computer from the paper tickets subsequent to call handling.

For many years direct dialling was only possible within limited localities, and later still only within countries. As direct dialling capabilities increased so did the number of operator assisted calls, especially over longer distances. The long distance international telephone call frequently requires operator assistance to provide charging information, language assistance, etc. In such circumstances the operator needs considerable knowledge and skill in using the international and national systems, sometimes via chains of other operators.

## A.4 FUNCTIONS OF THE TELEPHONE OPERATOR

In the earliest exchanges telephone operators were necessary to make the appropriate electrical connections between the 'calling' and the 'called' subscriber. With the advent and development of the automatic exchange an increasing number of simple calls could be dialled directly without intervention by an operator. Nevertheless, a significant number of operator assisted calls continued to be required for a variety of reasons. Examples included non-availability of direct dialling, codes unknown by subscriber or previous unsuccessful attempts at dialling direct. Such problems arose particularly when using the international network, where many countries could not be dialled direct and where the codes and/or dialling procedures were not known to the caller. However, as direct dialling was extended an increasing proportion of the operator handled calls required special facilities. Most of these facilities required the use of the operator as an intermediary between the calling subscriber and the exchange (e.g. to charge the call to a credit card), between the called subscriber and the exchange (e.g. to indicate acceptance of a 'collect call' - otherwise known as a 'reversed charge call') or between the calling subscriber and others necessary to reach a specified individual in person-to-person calls, (e.g. secretaries).

Thus far, the discussion has been exclusively related to demand calls, which are instigated by the calling subscriber and handled immediately. Another class of call, termed delay calls, are those which are for one reason or another stored in the exchange until a later time, at which both calling and called subscribers will be contacted. Such calls are common in international exchanges if the operator must go through a chain of other operators, a lengthy process during which connection to the calling subscriber is not required and, therefore, the operator will frequently offer to 'call back' when the outgoing connection has been established. Also for time zone differences and for unsuccessful call attempts the subscriber may wish to 'book' a call for a later time.

Another reason for using operators while making otherwise straightforward calls is to obtain charging information in advance of a call, to request price advice following a call or to book a fixed duration call.

Thus, the human operator not only enables subscribers to reach telephones not accessible by direct dialling, but enables them to receive language assistance, charging and price advice, additional charging functions (person-to-person, credit card, reversed charges etc.) and re-attempt administration. All of these facilities benefit from, or directly require, the use of a human intermediary between the subscribers, the exchange and the telephone network.

## A.5 THE NEED FOR NEW OPERATOR EQUIPMENT AND PROCEDURES

It can easily be appreciated that the operator's language skills, telephone system knowledge and access to special procedures were, and are, all useful in handling these calls. However, the forms of exchange existing at the time of this study placed a heavy procedural and administrative workload onto operators, particularly related to the use of paper tickets and the consequent need to re-key telephone numbers during successive call attempts. Moreover, there was a considerable lack of flexibility in assigning operator consoles to operator groups (eg. language groups). On the administrative side considerable numbers of paper tickets had to be handled and keypunched into a computer which kept records and produced subscriber invoices. This was a lengthy, expensive and error-prone task.

The L.M. Ericsson Company considered that, despite predicted technological advances (eg. speech recognition by computer) and further expansion of direct dialling, operators would be needed to assist in handling international calls for the foreseeable future. The Company also recognised that the new forms of automatic exchanges incorporating computer technology to provide stored program control would enable them to build new forms of operator console which would overcome many of the difficulties outlined above.

It was envisaged that the new consoles would incorporate computer terminals consisting of a datascreen or vdu and a keyboard connected to the electronic exchange via local computers, each serving up to 16 consoles. Information typed into the local computer and displayed on the datascreen would be transmitted to the exchange on the request of operators. Incoming demand calls to the exchanges would be identified and routed to appropriate 'free'operators who would then establish an auditory connection with the caller. As the subscriber described his requests the operator would enter details via the keyboard into the local processor, transmitting it to the exchange computers as required. The information to be added to the displayed ticket would include the necessary telephone numbers, names and call details. Once the call had been adequately described the operator would then initiate a call attempt by pressing appropriate keys. The computer would read the telphone number from the electronic ticket and dial towards it. The operator could then free the call from her position and be ready to receive another call. The computer would continue to monitor the call automatically, beginning and ending charging as required and storing this information for later processing on the accounts computer. If the call was unsuccessful, because the called subscriber was engaged or did not answer, another appropriate operator would be contacted and the relevant ticket would appear on her datascreen for her to interrogate the calling subscriber to decide how to proceed. If the caller wished the operator to try again later, the operator could enter an appropriate time or delay period into the relevant section of the ticket before disengaging the line to the caller and 'freeing' the ticket. The ticket would join the delay-call queue which would be periodically scanned by the computer and a free operator contacted at the appropriate time. To re-attempt the call the operator would simply press several function keys, not needing to re-enter the telephone number as in traditional exchanges.

In this way the details of a call and its subsequent history would be detailed on an electronic ticket which would be stored and processed by the computer, to be subsequently presented to any appropriate operator who could use it to provide advice or information to subscribers or to initiate the appropriate switching operations. The storage and processing of charging information would, it was hoped, be greatly hastened, simplified and made accurate.

## A.6 DESIGN CONSTRAINTS

The concept for the new operator equipment described in the previous section (A.5) had been taken to a preliminary design stage before the author was consulted. Consequently, it is necessary to describe the constraints imposed on the design from the beginning.

It had been decided that the operator would be seated at the terminal throughout and that no documents would be necessary. The datascreen would consist of a standard television monitor, but only 312 of the 625 raster lines would be used since the second set of interlaced scans would be unused.

Each raster line would be capable of displaying up to 600 independently addressed dots using a Programmable Read Only Memory (PROM) to generate the alphanumeric characters on a 7 x 9 matrix each matrix being formed on 9 raster lines. It was also specified that an easily available phosphor coating would need to be used. It had also been determined that the screen would be capable of displaying 16 lines of 64 characters each.

The keyboard requirements were considered to warrant a special purpose keyboard since it was felt that no existing off-the-shelf design would suffice. Four sets of keys were required, one set to enter digits, one set to enter alphabetic text, one set to move the cursor and edit the information and one set to control switching operations at the terminal and inside the exchange.

The configuration of the computer system required the provision of local computer processors linked to groups of up to 16 consoles, but each console would be capable of being assigned to any functional group (eg. operator language groups) irrespective of the local processor.

L.M. Ericsson wished to obtain an acceptable design for the chair and for the desk surfaces on which the datascreen and keyboard would be mounted even though the company did not manufacture furniture. Initially it was required that such a desk would enable the datascreen and the keyboard to be adjusted independently in height and perhaps in angle. A fully adjustable chair design was also required.

## A.7 PROJECT BRIEF

Broadly stated, the project brief was to produce ergonomics guidelines for the design of a computer assisted traffic handling position. The traffic handling position would consist of a datascreen and keyboard linked together to form a computer terminal and mounted on a console. LME were to manufacture the stored program exchange, the local processors supporting the terminals and parts of the terminals themselves, buying certain components from other suppliers as required. However, from the beginning it was recognised that it would be necessary to examine a series of related issues, some of which would not be of direct concern to an equipment manufacturer such as LME. Thus, for instance, it was known that the effectiveness of a datascreen would be influenced by the design of the ambient lighting conditions under which it was used. For this reason the project was required to consider not only the design of the datascreen terminal, its console and the call handling dialogue, but also lighting and work organisation.

## A.8 PROJECT PLAN

At the inception of the project it was suggested that the work might be conducted in four independent but successive phases.

Phase I: To examine the literature in order to establish design guidelines and plan later investigations.

Phase II: To conduct limited experimentation on separate components of the design as considered necessary as a result of Phase I.

Phase III: To test the entire operator position and call handling dialogue in prototype form under realistic environmental and workload conditions.

Phase IV: To carry out field trials in an early installation.

As is clear from section A.9, Phases I-III have now been completed and form the basis for this thesis. There have been some difficulties in mounting Phase IV but it is still hoped that it will take place.

## A.9 STRUCTURE OF REPORT

The project programme took place in three distinct phases which are reported separately in Chapters B, C and D. Within each chapter there are numbered sections to which cross references may be made.

Chapter B reports the results of Phase I, a literature review conducted in 1976, to ascertain from published documents which were the most important ergonomics factors for the design of the operator position and to describe the latest state-of-the-art. In the event the author was able to provide the design engineers with quite detailed recommendations for an ergonomics design specification.

Chapter C reports the results of Phase II, a series of preliminary experiments considered essential to resolve important conflicts between ergonomics recommendations, and to provide objective evidence where none could be found in the literature.

Chapter D describes the methods, results and conclusions of Phase III, an evaluative study of the entire operator position manned by operators and fed with a variety of simulated calls.

The final chapter (Chapter E) concludes the thesis by summarising the benefits of integrating ergonomics into the design process.

The references, appendices, tables and figures have been bound in the accompanying Volume II to enable them to be consulted in conjunction with the main body of the thesis, as it is read.

# CHAPTER B - PHASE I LITERATURE REVIEW

## B.1. PREFACE

The majority of this chapter was first prepared as a report for L.M.E. entitled "The human factors of ANE 403" (ANE 403 is the company code name for the entire operator position linked to the ANE family of automatic stored program controlled exchanges). The purpose of the report was to summarize the available literature relevant to the design and implementation of the new operator position. The contract specified that the literature review would "concentrate primarily upon the information task demands" but, where appropriate, would include "physical factors which would interact with these task demands".

The primary aim of the literature review report was to present the vital issues in a condensed and easily digestible form. The sponsor had not only requested a brief report but also one in which ergonomically acceptable solutions were recommended wherever possible. Published recommendations were frequently imprecise and there were major disagreements between authors. These disagreements could sometimes be traced to differences in experimental method, although this was not always fully described, but often they arose as a result of different practice in different fields of application. In order to provide L.M.E. with an acceptable total design solution the author of this thesis not only quoted the range of solutions quoted by individual authors but also tried to recommend narrower ranges or even single values for each design element.

Thus, a precise design specification was given taking into account the requirements of the task being performed and the sum total of the other recommendations. Where conflicts arose between specific recommendations, which could not be resolved by considering the relative importance of factors, and which were considered to be vital to system performance, experiments were suggested. The most important experiments were carried out and form the basis of Chapter C.

The chapter was in eight sections, the first four (B2-5) dealing respectively with datascreen, keyboard, dialogue and console, all of which were of immediate interest to the manufacturer. The next three sections (B6-8) dealt respectively with lighting, work levels and rest periods and the organization of working groups; issues which were ultimately the responsibility of the telephone administration using the equipment. However, since it was felt that each of these three latter issues interacted with those of equipment and procedure design, they were included and were ultimately used to instruct purchasing administrations in the most effective ways to install and use the equipment. The final section of this chapter comprised a detailed description of the recommended system design in summary form for easy reference.

It should be emphasised that this chapter was written originally in the first half of 1976. It has been left in this form because the information it contains formed the basis of the ergonomics recommendations and subsequent experimentation. Since this review was prepared, a great deal of publicity has been given to the ergonomic, radiological and task design aspects of vdu's. However, neither design guidelines as Cakir et al (1979), nor specific investigations, such as those reported in Grandjean and Vigliani (1980) fundamentally alter the correctness of the decisions made in this chapter or the thesis as a whole.

## B.2. DATASCREEN

The use of the electronic datascreen as a source of information for the operator of the new equipment made the visual components of the job much more important than in the operation of traditional exchanges. Well designed display hardware, logical dialogues and clearly laid out formats (i.e. tickets) were known to be likely to improve system performance and to reduce operator strain. The dialogue and ticket layout factors are discussed in section 4 of this chapter while the present section is devoted to the physical characteristics of the screen, whether determined by hardware or by software.

### B.2.1. Character Size

The first consideration of the designer was to ensure that characters displayed on the datascreen would be clearly legible to the operators. The factors affecting legibility were considered to be size, spacing, design, clarity and contrast. Figure 1 shows the important physical dimensions of characters.

The character size was defined in terms of the visual angle it would subtend at the eye, thus ensuring that screen distance and angle of view were also considered. Since 1 minute of arc was generally accepted as the limit of normal resolution this defined the minimum stroke width, and several authors investigating vdu's had used this limit to define the minimum character height at 10 minutes of arc, assuming a stroke width to height ratio of 10:1 (Hemingway and Erikson, 1969). Riche and Kinney (1969) had found that increasing the character height to 16 minutes led to shorter reaction times. Giddings (1972) had provided evidence of optimum character heights of 18 minutes for words and 21.5 minutes for digits (see Fig. 2). In view of the recommendations of these and other authors it was recommended that all characters should subtend at least 18 and preferably 22 minutes of arc at the eye of the operator.

The relative sizes of other dimensions of vdu characters and their spacing from others (see Fig. 1) was a well researched field, but few authors had specified clearly the degrees of performance decrement which might be expected from designs which deviated from the recommended levels. In the literature it was usually stated that character width should be at least 0.75 of the character height (e.g. Soar, 1955), although several authors had specified a minimum ratio of 0.67 (e.g. Poole, 1966). As explained in Section A.6 it had already been determined that in the present application 64 characters per line were required. It was therefore considered that narrower profile characters were necessary and a ratio of 0.67 width/height was recommended. Much of the literature had also suggested that characters should be separated by spaces of dimensions equal to 0.5 of the character height (Giddings, 1973), although authors such as Kuehn (1969) had suggested a space as small as 0.1 of the character height. Again the restrictions of screen space and scan pattern demanded that for 64 characters per line a suboptimal character space of approximately 0.2 of character height was the maximum available and therefore this was suggested. Line spacing, or 'leading' as it was called in the printing industry, had been recommended to be equal to 1 character height (Giddings, 1970), although, again, some authors had suggested that narrower leading, 0.5

of character height for example, may be acceptable. In the present application the screen size was determined to a large degree by character width, and sufficient space was available for line spacing of up to 1.5 character heights if the screen was to be viewed perpendicular to its surface and if the recommended character sizes were used (see Screen Size, Section B.2.7). A line leading of 1 character height was recommended if it was feasible for the screen to be viewed perpendicular to its surface.

The width of the strokes making up the characters was another factor known to affect legibility. In general it had been recommended that stroke-width/character-height ratios should be in the range 1/10 - 1/6. Berger (1944) had found that narrower strokes were preferable for self luminous numbers on a dark background (see Section B.2.5 for discussion of direction of contrast), favouring stroke-width/character-height ratios of between 1/10 and 1/13. In view of the rather small character width/height ratios recommended above, and the consequently small areas which would be enclosed by the strokes in letters such as B and P, it was considered advisable to attempt to produce stroke widths with ratios between 1/10 and 1/8 of character height. Soar (1958) had found that variations of stroke widths within characters could improve legibility but it was thought that this technique had to be rejected on practical grounds for the present system and had been rejected by Berger (1944) for aesthetic reasons. The minimum acceptable stroke width was independently determined by another attribute of the system, the character generation technique. Because characters were to be generated from a 7 x 9 dot matrix on a raster scanned t.v. (see next Section, B.2.2) the dots had to overlap to produce the perception of continuous strokes. Spot sizes of 1.2 times the raster line space had been recommended for non-interlaced displays, while 1.7 times the raster line space had been recommended for interlaced displays in order to account for delays between regeneration of adjacent scans (Brown, 1967). In the present application, it had been determined that the interlace would not be utilized, the displayed raster lines would, by necessity, be spaced at distances equal to 1/9 of the character height, since each character would be made up of selective brightening of nine raster lines. If a spot size of 1.2 times the raster line spacing was used it was realised that this would result in strokes which would be only 1/7.5 of character height as a minimum. If brightness were to be increased it was noted that this spot size would tend to increase and may have easily exceeded the maximum acceptable stroke width of 1/6 of character height.

B.2.2. Character Generation Technique

As explained in the Introduction, it had been determined that the characters displayed on the ANE 403 datascreen were to be constructed on a 7 x 9 matrix of dots (7 dots wide x 9 dots high). Vertically the dots were to be defined by the t.v. raster scans passing through the character, while horizontally the dots were to be formed by selectively brightening parts of these raster scans. The literature was quite clear in recommending at least 10 raster scans per character and several experiments had shown that response time and errors decreased as more scans were used (Fig. 3). Elias et al. (1965) and Hemingway and Erikson (1969) had shown similar results, while Shurtleff (1967) had quoted performance figures for different sizes of character with varying numbers of raster scans passing through them. Unfortunately, these experimental results related most closely to televised transmission of printed text, and not to computer generated characters, so the results had to be treated with some caution. Gould (1968) had also pointed out that many of these experiments were related to the perception of individual characters whereas words could be legible with fewer raster lines per character (Elias et al. 1965; Shurtleff, 1967). As stated above, in the present application only 9 raster scans would pass through each character but characters were envisaged which would be larger in size than had been used in the experiments quoted above. It was felt that if the number of raster scans per character could have been increased, to at least 10, this would have been advantageous but that a dot matrix having 9 vertical dots was considered adequate.

A technique had been developed for domestic television text systems which provided intermediate dots between those produced by the PROM (Programmable Read Only Memory). The technique, called "Character Rounding" (Kinghorn, 1975), interleaved intermediate dots on each raster scan and produced dots for the otherwise unused interlaced raster scans by the application of simple rules. Better formed characters were provided by this technique without greatly adding to the cost (see Fig. 4). An additional benefit of character rounding in the present application was that it would be possible for the stroke width of characters to be smaller than for corresponding non-rounded characters for given brightness levels. This was explained by pointing out that given that Brown (1967) had recommended that the spot size of interlaced displays should be 1.7 times the inter-raster distance (see Section B.2.1) then it would have been possible for spot sizes and, consequently stroke widths, to be a minimum of 1/10.5 of character height, since each character would be formed on 18 raster lines. Unfortunately interlacing was considered to have two unwanted side effects, a slower actual refresh rate (25 Hz instead of 50 Hz) and a possibility that the image would appear to drift down the screen. It was recommended that the technique of character rounding be tried on the equipment proposed and that legibility experiments should be conducted to evaluate the usefulness of the technique.

### B.2.3. Character Design

It was realised that the design of letter and number shapes would be as important a factor in the legibility of electronic characters as it had been shown to be for printed characters. A number of authors had carried out experiments to compare different electronic character designs. Vartabedian (1971a) had demonstrated that 7 x 9 dot matrix symbols, drawn with circular dots, were superior to a variety of other designs on both reaction times and error rate. Vartabedian (1971b) had also shown that, in a search task, single words in 'upper case' (capitals) were scanned 13% faster than words consisting of 'lower case' letters. Vartabedian (1971c and 1973) had included in his work a complete ASCII (American Standard Code for Information Interchange) 96 character set for a 7x9 dot matrix which incorporated a shift of the matrix to produce adequate lower case characters. Huddlestone (1974) had compared an alternative 7x9 character set with those of Vartabedian

(1971c) and reported that his character set produced less reading errors than those of Vartabedian; but Huddlestone gave no explanation for this improved performance. Berger (1944) had discussed experiments on the form of numerals for car number plates and had included his recommended character set for numerals.

The aim of character set design was to improve the ease with which the letters and numbers could be distinguished from one another. Certain character pairs were recognised to be more easily confused than others (e.g. 8 and B, C and G). To reduce these confusions the properties distinguishing between confusable characters had to be highlighted (e.g. "B" could be given serifs to make it more distinguishable from 8). The character set originally proposed for ANE 403 had been examined and changes to certain characters were recommended (see Figs. 5A and 5B).

### B.2.4 Cursor Design

A cursor is a means of indicating to the operator the position on the screen at which keyed characters will appear. It takes its name from the hairline 'cursor' provided on such instruments as the slide rule.

The cursor to be used on the datascreen for the new operator position had to be easily seen, its position had to be clear and it was essential that it did not obscure any other displayed characters. Vartabedian (1970) had examined several cursor forms and rates of blinking in terms of search times and tracking performance. Vartabedian had found that a hollow box cursor, consisting of a rectangle enclosing the character space, and blinking at 3 Hz was the most effective, but that both box cursors and simpler underline cursors were equally favoured by subjects. In view of the narrow spacing which had been necessarily recommended for the datascreen characters (see Section A.2.1) it was felt that a box cursor design might have interfered with the legibility of certain character forms. An underline cursor flashing at 3 Hz was thus recommended for ANE 403. The "insert character" editing function was necessary to enable the operator to insert characters into existing text strings. As described in Section B.4.9 it was recommended that the new character space produced by the insert character command should appear to the left of the original cursor position and that the cursor should be positioned below the space to enable the insert character to be entered. To make the position of the inserted character clear it was recommended that the cursor should have a small vertical bar up from its left hand end. It was also thought that this bar would improve the ability of the eye to 'lock onto' its position as suggested by Bedwell and Stone (1976).

### B.2.5 Character/Screen Contrast and Character Brightness

In order for a character to be made distinguishable from the background upon which it would be displayed it was known to be necessary for there to be a brightness or colour contrast between them. Colour contrast was defined as resulting from differences in the wavelengths of light emitted from, or reflected by, the background and the character. Although colour contrast was considered effective for large areas it was thought to be less effective for character perception (MacLean, 1965) a finding attributed to the fact that the lens of the eye was known to be not colour corrected, and in consequence the focussing points on the retina of two widely separated colours were different. Brightness contrast resulted from a difference of luminance of the characters and the background, whether this was due to difference in self illumination (as a datascreen), or to the differential reflection of ambient light (as on the printed page).

The average luminance levels of printed matter had increased over the previous years as general illumination levels had risen. Gould (1968) had claimed that, at the time he wrote, under typical conditions white paper reflected around 50 mL (159 cd/m$^2$) and he had suggested that this level should be used as an estimate for symbols on a CRT display (Gould assumed light characters on a dark background). Gould had pointed out that the recommended luminance level of 159 cd/m$^2$ was only approximate, and that levels above 80 cd/m$^2$ were probably acceptable, assuming adequate contrast. Although at the time of the study many vdu tubes could reach luminance levels higher than 150 cd/m$^2$ it was considered that this figure would have to be obtainable at below maximum tube output, since it was assumed that at high beam currents tube life could be reduced and/or resolution made poorer.

It was pointed out that, if character brightness was to be used as a coding device (e.g. to distinguish newly entered text from that existing) the brightness levels chosen for these texts had to be clearly different from each other. A 20% difference in brightness level was considered sufficient (see Section B.4.6).

It was recommended that the luminance of the characters on the datascreen should be capable of reaching at least 150 cd/$m^2$ without unduly affecting tube life or resolution (as an example it was recommended that stroke width/character height ration should not exceed 1:6). If 150 cd/$m^2$ were to be the luminance of the brightest characters 120 cd/$m^2$ was recommended for the second brightness level (i.e. 20% difference). Ideally, it was felt to be beneficial if the newly entered and the existing characters could have luminance of 190 cd/$m^2$ ans 150 cd/$m^2$ respectively provided that tube life and resolution would not be unduly affected.

The eye was considered capably of detecting luminance differences between the character and the empty screen as small as 1% (Van Cott and Kinkade, 1972) this represented a contrast* of only 0.01, but it was felt that this figure bore no relation to the contrast considered necessary for accurate and easy reading. Unfortunately, many of the studies reviewed were concerned with threshold contrast levels and with limited response criteria (e.g. detection of blips on radar scopes). In a practical attempt to determine the optimal contrast between character and background for computer displays, Stocker (1964 and 1966) failed to provide a recommendation, because perceptual or reading speeds continued to increase as contrast far exceeded that obtainable on CRT's. Howell and Kraft (1959) used simulated CRT characters, and found that, when the contrast* had been increased

---

* The CIE (The International Lighting Commission) defines contrast as follows:

$$\text{Contrast} = \frac{L_o - L_B}{L_B}$$

Where $L_o$ is the object luminance and $L_B$ is the background luminance. Thus, when viewing a dark object against a light background as in typed script the contrast is negative (because $L_o < L_B$). Where authors expressed contrast as a luminance ratio ($L_o/L_B$ or $L_o:L_B$) this has been converted to contrast as defined by the CIE formula shown above.

from 12 to 37, there has been little increase in legibility, provided that the characters subtended greater than 16 minutes of arc at the eyes and were not judged as blurred. In general, Howell and Kraft (1959) recommended a contrast of 29, with 14 being "acceptable". In practice, contrasts of greater than 19 had rarely been obtained without the use of contrast enhancing devices. Because CRT symbols were considered to be more blurred than printed symbols, Gould (1968) had stated that considerably higher contrast was needed on CRT's than was typically found on hard copy. The actual contrast on CRT's could be improved by reducing the background luminance produced on the t.v. screen from within, by reducing the reflected ambient illumination from the whole screen or by increasing the luminance of the characters. Poor contrast was also reportedly less disruptive as character size was increased (Howell and Kraft, 1959). Decreasing the luminance of the background could, it was thought, only be achieved by improved hardware design (e.g. by using finer or thinner phosphors). Increasing the character brightness had a known tendency to produce more flicker, scattering and reflections, making the image less distinct and reducing any beneficial effects (Poole, 1966). The use of much brighter but at the same time 'clean' images was considered to await major breakthroughs in phosphor development. The most practical means of improving contrast appeared to be to reduce the degree to which the screen would reflect ambient light. The use of low-reflectance surfaces on the screen was thought likely to improve contrast, but unfortunately such surfaces were thought to affect character clarity, producing blurring of the image.

Although it was generally considered at the time of the report that contrast should be maximised it was also thought possible that a higher background luminance, and therefore somewhat lower contrast, might lead to less eye fatigue. This result was thought possible if the non-structured properties of the screen could make it difficult to obtain sufficient cues for focussing and binocular convergence. Fatigue was also thought to be potentially attributable to 'hunting' of these ocular mechanisms in a vain attempt to improve the perceived image quality. By increasing background luminance the face of the screen was thought by this theory to become a surface whose distance from the eyes might be more easily determined. It was recommended that such a hypothesis be tested experimentally.

The question of whether to use dark-on-light or light-on-dark characters was reported as usually dealt with under the heading "direction of contrast". Shurtleff (1967), in discussing this issue, had suggested that, for "intermediate" values of character and background brightness, the direction-of-contrast was considered as probably not a major factor in legibility since the largest differences reported had been 4% at these levels, (Seibert et al. 1959 and Kelly, 1960). At extremely high levels of illumination, where dark-on-light symbols were shown to be superior to light-on-dark, overall accuracy of identification was poor and the condition was considered unacceptable for system application (Shurtleff, 1967). The larger area of bright screen found on dark-on-light, or 'reverse video' as it was sometimes called, was more likely to give rise to flicker perception. <u>It was therefore recommended that 'light-on-dark'</u>

positive contrast was used in the present application.

It was recommended that a contrast level of 15 should be possible but that 10 would prove acceptable if reflected specular reflections were avoided. It was recommended that this contrast should be attainable under 300 lux illumination, without causing poor character definition or reducing the screen life appreciably (300 lux is recommended in Section B.6.1).

B.2.6. The Use of Filters and Meshes to Enhance Contrast

Neutral density filters placed in front of the screen were known to reduce the luminance of the characters proportional to the filters' transmission coefficient, whereas ambient light reflected from the screen would pass through the filter twice and consequently its intensity would be reduced by the square of the transmission coefficient, thus markedly enhancing the contrast.

Lally (1966) had described a filter which transmitted less light towards the display than away from it, a technique which it was thought would further enhance contrast.

Circular polarizing filters had the advantage of totally eliminating light diffusely reflected and specularly reflected from the screen. Unfortunately, many of the polarized filters were known to have polished surfaces which caused their own specular reflections (e.g. of operator's white shirt). Hyman, (1967) after reviewing a variety of contrast enhancing devices, had recommended the use of a circular polarizing together with an anti-reflective coating, but it had to be remembered that such coatings diffused the character luminance still further. It was known that such diffusion could be reduced somewhat if the filter was placed very close to the screen. Giddings (1971) had classified a variety of optical filters and evaluated a number of them. After rejecting polarizing filters as "giving too little improvement" he had continued to compare 4 filters; a fibre optic faceplate, two types of mesh and a thin film multi-mesh filter, called 'Hycon'. The latter consisted of 14 aligned layers of vacuum deposited metal inside a glass sandwich, making its transmission properties very directional but quite high. The Hycon filter had given better performance measures on the task, with coarse and fine meshes ranking second and third. Subjectively, however, the Hycon had been found to be the least acceptable filter since its mesh structure had interfered with the resolution of the screen and there had been specular reflection from its front face. Giddings (op. cit) had found that meshes, although giving poor performance, were considered more pleasant to view and did not restrict posture to the same extent as the Hycon filter which was very directional. The 3M company made various louvred faceplates (light control film) with louvres oriented to reduce the light levels reflected from the screen but again they were thought to give rise to specular reflection from their front face (a light matt finish was available which was claimed to reduce specular reflections without unacceptable loss of image clarity but it had not been evaluated by any independent research body). The Hycon and 3M films had an added disadvantage, that they were likely to restrict viewing angles severely, necessitating much more careful console design, and possibly causing postural difficulties.

Knowles and Wulfeck (1972) had compared four types of CRT under very high illumination levels for possible use as datascreens in high altitude flight. In these experiments three of the CRT's had had a different contrast enhancing treatment while one was untreated. The most effective treatment had been shown to involve the use of a special light-filter bonded to the face of the CRT. This filter had a narrow band of transmittance designed to match the spectral output of the CRT to the peak of the spectral sensitivity of the eye, and to greatly attenuate light from other regions of the spectrum. This filter had performed extremely well, having a reflectance of 0.21% measured at 90° to a screen illuminated by 10,000 ft. c. (929 lux) at 60° to the screen. A thin film phosphor used on another CRT had given a reflectance figure of 0.83% under the same levels of illumination while the untreated screen had reflected 73% of incident light. Knowles and Wulfeck had stated that the visual performance in their task was primarily a function of background brightness and therefore reflectance. Contrast provided by a low reflectance screen was reported to be most important at high ambient light levels whose effect would otherwise be to increase background luminance significantly.

Since there had been conflicting reports on the value of contrast enhancing filters and anti-reflective treatments it was recommended that an experimental investigation of such devices was conducted. (There is a further discussion of filters and their method of operation in Chapter C and in Appendices 6 and 7).

## B.2.7. Screen Size

The size of the datascreen was considered to be best described in terms of the angle which it would subtend at the operator's eye. As described in Chapter A, the ANE 403 specification required the display of a 64x16 matrix of characters, together with spaces and line leading (Section A.6). Assuming that each character height would subtend 22 minutes of arc at the operator's eye, as recommended earlier (Section B.2.1.), this gave the following screen sizes for three possible viewing distances (500mm, 700mm and 1,000mm)

| Viewing distance | Character size (mm) | Character space (mm) | Screen dimensions (mm)* (W x H) | Diagonal Screen size (in)ǂ viewing | actual |
|---|---|---|---|---|---|
| 500 mm | 3.2 x 2.1 | 0.64 | 175 x 125 | 8.5" | 10" |
| 700 mm | 4.5 x 3.0 | 0.9 | 250 x 178 | 12.1" | 14" |
| 1000 mm | 6.4 x 4.3 | 1.28 | 357 x 255 | 17.3" | 19" |

* The screen height quoted was not necessary, given the line spacing recommended, but anticipated the use of a standard tube with a 4:3 width to height ratio.

ǂ These sizes of screen were obtained by adding a margin to the viewing area and rounding up. The sizes quoted were not known to be commercially available and a larger size was recommended if those quoted were found to be unobtainable.

The sizes specified above related to datascreens placed perpendicular to the line of sight. Screens angled away from the perpendicular to the line of sight were recommended to have the relevant character dimensions adjusted, but it was thought possible that such designs could lead to accommodation difficulties, because parts of the screen would be further away than others, requiring the user to refocus when looking from section to section of the display.

The solid angle subtended at the eye by the screen was known not only to affect the perception of characters of a given height but also to affect the perception of flicker. The larger the area which would be subtended at the eye by the screen (and therefore by the character) the greater would be the accuracy with which the individual characters could be read, but also the greater would be the likelihood that flicker would be perceived for a given refresh rate and screen brightness (see next Section B.2.8.). Also beyond a certain point the readability of words would be impaired. All the screens listed above, if viewed at the recommended viewing distances, were likely to lead to approximately the same amount of perceived flicker since they all would subtend roughly the same angle at the eye (i.e. approximately $27^{o}$ subtended by the diagonal of each screen).

It was recommended that the screen diagonal should subtend approximately $27^{o}$ of visual angle at the eye of the operator.

B.2.8. Phosphor Characteristics

Phosphors used in cathode ray tube displays are doped crystalline elements which fluoresce during electron bombardment and phosphoresce after bombardment has been removed. Many of the visual characteristics of a datascreen display are determined by the dynamic physical characteristics of the phosphor employed (e.g. colour, brightness, decay rate, etc.).

A wide variety of phosphors were available, each with different properties. The choice of phosphor had to be determined by the application to which it would be put. Knowing that the physical characteristics, of colour, persistence and burn resistance, were not independent of each other, it was apparent that the choice would inevitably involve resolving conflicts between several requirements. The ideal phosphor was thought to be one which would provide good resolution, reduce flicker to a minimum, while prolonging tube life, and providing a high perceived image brightness for reasonable beam current flow (the latter was considered important if the heat generated by the tube proved to be large and had to be minimised).

The required spectral output of the phosphor was difficult to specify because colour perception was known to be highly subjective and the characteristics of available phosphors were problematic because there were known to be noticeable variations between the same named phosphors from different manufacturers (Bell, 1970). Spectral output was defined as the relative radiant energy emitted at various wavelengths, but phosphors having different spectral outputs could have the same apparent colour. The human eye was known to be sensitive only to a narrow band of electromagnetic radiation and, when light adapted, to be maximally light sensitive in the yellow-green sector of the electromagnetic spectrum (around 555 nanometers). Phosphors which emitted substantial intensities of light close to 555 nanometers were therefore considered "more acceptable", (Graham, 1965). Figure 6 shows a comparison between the spectral sensitivity of the eye and the relative luminous output of Phosphor P39 which shows how close the spectral output of this phosphor matched the sensitivity curve, indicating an efficient energy transfer was likely and suggesting that this phosphor should be considered, at least on this criterion.

The persistence of a phosphor was defined as the time required for an image to decay to a given fraction of its initial intensity. The 10% and 1% levels were usually selected for reference, with the former taken as the 'minimum perceptible brightness in high ambient lighting', and the latter as the 'approximate limit in dimly lighted conditions'. The persistence curves quoted by such bodies as the American Joint Electron Device Engineering Council (JEDEC) represented the original registration data but it was known that many of the initial performance figures for a phosphor degraded markedly over their life-time (e.g. Hayman, 1969).

It was known that when a CRT was not refreshed sufficiently frequently it appeared to flicker. The relevant behaviour measure of flicker was the "critical flicker frequency" which was conventionally defined as the level below which flicker was perceived by 50% of any given sample. Peripheral areas of the visual field were known to be more sensitive to large areas of flicker than the central area or fovea and in a large room flicker could be perceived more easily from screens seen in peripheral vision. It was also known that the critical flicker frequency tended to decrease with viewer age, and logarithmically with increase in the angular size of the source, or the luminance of the source. In a CRT longer persistence phosphors were considered to give rise to less flicker perception at a given regeneration rate (see Gould, 1968). Unfortunately longer persistence phosphors tended to 'burn' more easily (see below) and gave rise to 'trails' from fast moving images (such as cursors).

Phosphor efficiency, in converting electron energy to light energy, was known to be typically less than 10% and large heat dissipation was usually necessary to prevent over-heating. It was anticipated that excessively bright traces or long dwell times were likely to lead to local burning of the phosphor, causing discolouration and reduction of the phosphor efficiency. Standard phosphors could be classified by burn resistance, and, where intense, sharply focussed immobile spots were required, a high burn resistance phosphor was usually specified. It was recommended that images likely to burn the phosphor were to be reduced in intensity and/or periodically moved slightly to distribute phosphor wear evenly.

It had been specified that the screen used in ANE 403 would be used for many hours each day and would contain fixed fields with constantly displayed labels. To provide adequate contrast in the ambient light levels recommended in a later section (Section B.6), it was known that the brightness of characters would need to be high. A refresh rate of 50 Hz was the most common found in Europe at that time and this was considered acceptable for a number of phosphors, only giving rise to appreciable flicker when most of the raster lines were displayed at high brightness (e.g. using reverse video) and when the screen was close to the observer. The phosphor P39 seemed ideal in many respects for use on ANE 403, but its efficiency and burn resistance were clearly lower than those of another contender, the standard P31 phosphor. Hayman (1969) had stated that P31 phosphor should have been refreshed at 55 Hz or faster and that there was a danger that significant flicker could be perceived at 50 Hz. If significant problems of flicker occurred with P31 then it was recommended that P39 phosphor be used.

## B.3. KEYBOARD.

As explained in Section A.3.3. it was decided that the telephone operators' means of controlling the exchange and of entering data onto the screen would be via a number of sets of keys on a keyboard. A great deal had been written about keyboards during the previous 20 years but most of it had related to very general purpose keyboards, to the physical characteristics of the keys or to specific applications, (Remington and Rogers, 1969; Hillix and Coburn, 1961).

### B.3.1. Physical Characteristics of Keys

The physical characteristics of the push button keys of a keyboard (see Fig. 7) were known sometimes to affect the speed and accuracy of performance. Considering the many possible combinations of key size, shape, force and displacement and the likely interaction of these variables with task, it was surprising that few factorial experiments had been conducted.

Deininger (1960) had conducted the best known and probably the most extensive research on the effects of key characteristics on performance. His research had suggested the optimum size, space, displacement, force and feedback for ten-key telephone entry keysets. Deininger (op.cit) had found that the largest performance increase was found by increasing the button-top size from 9.5 mm to 12.7 mm, which had reduced keying times in his experimental task from 6.35 sec. to 5.83 sec. and errors from 7.3% to 1.3%. Variations of depression force, between 100 and 400 gf (approx. 1-4N), and variations of maximum displacement between 0.7 mm and

4.8 mm, had produced insignificant differences in subjects' performance (note that these were the total ranges used). Subjects' preferences, however, had been for lighter touch keys with shorter operating displacement. Kinkade and Gonzalez (1969) had used eight experienced typists to evaluate the influence of key force and displacement and had found that performance was best when force and displacement were at relatively low levels. They had recommended key forces between 25 and 150 grams (approx. 0.25-1.5N) and key displacements between 1.3 mm and 6.4 mm.

It was stated that two commercially available keyboard systems, "Rafi Contactlose Tastaturen RC 72" and "Tastaturen RS 74" had specifications which lay within the ranges specified by the majority of the literature and were therefore equally acceptable on these criteria.

| | Recommended | RC 72 | RS 74 |
|---|---|---|---|
| Key top size (mm square) | 12.7 | 12.7 | 12.7 |
| Key spacing of centres (mm) | 19.05 | 19.1 | 19.1 |
| Displacement (mm) | 1.3-6.4 | 4.5 | 2.5 |
| Actuating force (Newtons) | 0.25-1.5 | 0.7 | 0.7 |

B.3.2. Key Feedback

It was reported that in normal typing and other key-pressing tasks there is kinaesthetic (muscle and joint) and tactile (touch) feedback from depressing the key, auditory feedback from the key press and/or the print mechanism, and finally there is visual feedback from the keyboard and from the output display.

Although general psychological literature on learning had suggested the necessity of feedback for efficient keying performance a review of the literature on keyboards indicated that a great deal of feedback was unnecessary for maintaining skilled performance after learning had occurred. Deininger (1960) had found no statistical differences in speed and error performance when auditory and kinaesthetic feedback were added to push buttons with 3.2 mm travel and 0.97 N displacement force. Moreover Diehl and Seibel (1962), using experienced secretaries and stenographers had found no performance decrement if visual cues, auditory cues or both types of feedback were masked. Although Galitz (1965) had found that the absence of auditory feedback 'bothered' typists, no typing performance figures had been given. Kinkade and Gonzalez (1969) had found that the addition of kinaesthetic/tactile "snap" action feedback had not affected speed and, in fact, had led to more errors being made. Other key feedback features such as bottoming, breakout force etc. were thought to influence key-pressing performance but they had not been studied in detail.

On the ANE 403 there were to be four sets of keys with very different functions, some of which were considered likely to benefit from some form of artificial feedback. The Rafi keyboards proposed, worked by the so-called 'Hall effect', containing magnetic switches which involved no mechanical actuation thus providing no inherent cues as to when the key operated until the bottom of the stroke was reached. During skilled data entry via the alphabetic and numerical keyboards feedback additional to that provided by the keys was not thought to be essential and, that, providing the system response was fast (i.e. less than 0.1s), the operator would be able to obtain visual feedback from the datascreen that the keys were working.

However, it was considered essential that keys with important system functions (e.g. speak to calling subscriber) and for which the operator would receive no immediate indication, either visually from the datascreen or auditorily from the headset, might benefit from the addition of additional feedback from that provided by the keys themselves. This was considered especially important if multiple keypresses would change the system response (e.g. a second press of the "Speech Monitor" key would nullify the first press). It was explained that if the system response could be delayed for greater than 2 sec. some form of instantaneous feedback would be beneficial. The keys required to control the exchange directly, causing switching operations, were thought likely to benefit from the addition of feedback if the tones to be heard over the headset would be insufficient to indicate system state immediately. The editing function "character insert" was not going to give a system response until the operator attempted to enter characters and, again, additional feedback was considered to be beneficial.

In some cases auditory feedback, in the form of a click from a relay or a short tone, was considered sufficient to indicate that a code had been transmitted to the system, but in other cases it was thought advantageous if the key itself lighted to indicate the state selected. It was also suggested that alternatively an easily distinguished message, code or shape could be included on the datascreen.

Some forms of feedback additional to that provided by the keys themselves were considered likely to be beneficial for certain function keys and for the "character insert" key.

B.3.3. Colour of Keys

It was reported that in general the colour of the keys was not an important factor, except insofar as that it would be used by a "hunt and peck" operator alternately looking at the screen and the keyboard. Dark keys with light lettering were reported to be preferable to light keys and dark lettering and it was stated that these had the added advantage of not being as distracting if reflected in the screen. However, colour was thought to be a useful means of coding the function of sets of keys and of providing additional cues to indicate the functions of adjacent keys. A scheme was suggested for the use of colour coding on the keyboard and this is shown in Fig. 11.

B.3.4. Means of Repeating Cursor Movements

A means of controlling continuous movements of the cursor was described to be useful for ANE 403 system to reduce the need for multiple key presses. One technique commonly used at the time to provide such a 'typomatic' function was to start the continuous action of the cursor after a key had been held down for greater than 0.5 sec. Another technique, which it was considered might be equally easy to use, involved the use of a two-position key such that slightly greater key pressure would cause the typomatic function to be executed.

The 'typomatic' cursor action was recommended to be useful for moving along lines of text to locate specific characters but was not recommended for line feeds up or down since, in the light of the authors experience, such typomatic functions would not be used over small numbers of lines such as recommended in Section B.4.2.

### B.3.5. Keyboard Slope

There was very little literature found relating to the question of typing performance as a function of keyboard slope. A review of literature on fatigue, by Moneta (1960), had reported that British typewriters appeared to have a standard slope of 30°. This angle seemed to have been based on convention rather than empirical data however, since Lundervold (1951) had shown that it was difficult to hold the forearms at this angle without effort. Moneta had therefore suggested that a slope of less than 30° would be preferable but had not indicated a particular value. Scales and Chapanis (1954) had examined keyboards with slopes from 0°-40°, finding no significant performance differences but a majority subjective preference for angles between 15° and 25°. Dreyfuss (1959) in his tables of anthropometric data had indicated an 11° slope as optimum and 20° as maximum permissible. Galitz (1965) investigating the effects of slope on typing performance had studied slopes of 9°, 21° and 33°, found that 21° was subjectively preferred. Galitz had concluded that keyboard slope should be adjustable between 10° and 35°, to account for subjects preference, since there was little performance difference within this range.

A fixed keyboard angle of approximately 20° was recommended.

B.3.6. Thickness of Keyboard

It was known at the beginning of the study that the keyboard had to be mounted in a box which would not only hold it at the required angle (see Section B.3.5.) but would house the switches of the keys and a certain amount of bulky electronics. Moreover, it was considered essential for the keyboard box to be a separate unit from the datascreen.

It was generally held that a keyboard should be placed at a height to enable the users to hold their lower arms horizontally when keying. However, if the arms were to be held in this position the distance between the hands and the top of the thighs would be rather short and, if sufficient thigh clearance were to be allowed, the space for the desk thickness and keyboard would be limited (see Section B.5.2.). Thus it was recommended that the front of the keyboard be as thin as possible and that the physical depth of the push buttons constituting the keyboard, including the activating mechanisms and electrical contacts, be as small as possible to enable the keyboard to be mounted at the correct height while allowing maximum leg room under the desk. One of the Rafi keyboards (Contactlose Tastaturen RC 72) had key tops more than 38 mm above the lowest point of the key assembly so that the top of the keyboard had to be at least 40 mm above the work surface. The top of the other Rafi keyboard (Tastaturen RS 74) only needed to be 15 mm above the work surface and was therefore more acceptable on this criterion.

## B.4. DIALOGUE DESIGN

The term dialogue was used, in the context of ANE 403, to refer to the information flow which would occur between the telephone operator and the computer via the datascreen and its associated keyboards. Since each operator position was likely to handle a great many telephone calls, small reductions in error rates or in the speed of the transactions were predicted to have significant economic benefits. A typical dialogue sequence was described and specific design recommendations were made, relating to format, keyboard layout and procedure.

### B.4.1 Typical Dialogue Sequence

Operator Signature. Logging onto the terminal would be a straightforward procedure and, since it would only be done once per shift, it warranted less attention than other tasks. It was considered that the power switch could possibly be incorporated into the headset socket, but such a consideration was likely to be of minimal importance. If an operator entered an inccrrect or unacceptable code the message subsequently appearing in the ORD I (abbreviation for ORDER 1) field was recommended to be more explicit than a simple repeat of the request "enter signature" (e.g. "signature unacceptable"). Once logging on had been accomplished, the signature would appear on all tickets being deal with and would be permanently attached to tickets originally sent to that operator (in the field labelled SBO, an abbreviation for signature of backing operator) and those last dealt with by her (in the field labelled SCO, an abbreviation for signature of completing operator). The signature fields would be only infrequently used by operators dealing with tickets, and consequently it was recommended that they could be placed in less accessible places on the screen.

Arrival of demand call. Once the operators had "logged-on" the system would allocate calls to them according to the workload of their work group. Incoming demand calls would first be signalled at an operator's position by the display of a ticket which could be partially completed. It was thought that the operator should have a little time (e.g. 3-5 s). to appreciate the nature of the call and to ascertain whether it had been correctly routed to her before she accepted it, (perhaps by pressing the button for speech monitor (SM) or a separate "ACCEPT" button). The system would immediately place the calling subscriber (sometimes called the 'A' subscriber), or an operator in another country, into voice contact with the operator: sometimes after giving a pre-recorded message. If the call had been incorrectly routed, as ascertained from subsequent conversation with the calling 'A' subscriber, a simple and fast transfer procedure was recommended. If the caller had been correctly routed, the operator would then wish to obtain certain items of information necessary for call handling and billing. The form of dialogue, recommended for ANE 403, involved the completion of a pre-formatted ticket, displayed on the datascreen, information items being entered into pre-existing labelled fields. By the use of cursor movements the operator would access any of the open fields and enter information. This system had the advantage that fields would be coded by position as well as label, and it maintained the layout of tickets throughout the procedures. If, instead, the fields for completion by the operator had been scattered, and consequently demanded complex cursor movements to reach them, special codes would have been required to cause the cursor to jump to fields. These special codes could have been entered on the alphabetic keyboard or entered from a special set of function keys individually marked with the field names. A more practical solution seemed to be to design the ticket such that the 'open' fields were near one another and that cursor movements were simplified, and therefore this was the technique recommended (see Section B.4.2. and Fig. 9).

Ticket completion. The operator would enter the telephone numbers of the calling 'A' and called 'B' subscribers in the A-NR and B-NR fields respectively and the name of the person required in the PER field (if a person-to-person call was required). If the complete telephone codes for the called subscriber were not known the operator would be allowed to enter the name of the country and/or city and, if this name was subsequently 'recognized' by the system, the code would be entered automatically, the name being taken to the end of the number in the field for later reference if required. If a name was not recognized by the system its status would be clear (e.g. by remaining brighter than the accepted text) and a message in the field reserved for messages from the computer (ORD I) would indicate to the operator that the number had to be found by external reference (e.g. inquiry service or directory). The operator would be required to enter certain information about the category of the call, since this would determine its pricing and/or priority. Coinbox calls, requests for price advice etc. were to be dealt with using the same technique. The call categories would be entered into the field reserved for commands to be entered into the computer (ORD II, abbreviation for ORDER 2) individually and entered into the system utilizing the key marked TRANS. As these were examined and accepted by the system the relevant codes would appear in the category (CAT) field. The disadvantage of this technique was that codes would have to be entered one at a time rather than in groups, but it did ensure that each category would be accepted before the next was entered. Providing the acceptance procedure would take less than 2 seconds from pressing the key marked TRANS (for transmit) to its acknowledgement in the CAT field, this technique seemed satisfactory.

Call Initiation. Once the necessary information had been added to the ticket via the command input field, labelled ORD II, the operator would initiate a call attempt by pressing the button marked 'B' (for the 'B' or called subscriber). When the operator was satisfied that the system was making the call he could press the Time Start (TS) button and the button marked FREE so that he would be in a position to accept the next call. The system would monitor the line, and begin timing and charging the call once the called ('B') subscriber answered. If 'B' did not answer, was engaged, or if the line was congested, a free operator would be contacted to handle the call. It was recommended that the reason for contacting an operator should be made clear by computer completion of three fields, the 'Remarks' field (REM), the field indicating the status of the calling (A) and called (B) subscribers, and finally in the request from the system written in the ORD I field.

Booking a Delay Call. If a delay call was requested by the 'A' subscriber, the time requested could be entered into the system. The computer would check the time intervals around those requested and, if the workload of booked calls was sufficiently low, it would accept the booking. If the booking could not be accepted, an alternative time or times would be suggested. One facility, which was considered to be useful in booking delay calls, was to enable operators to enter periods during which a delay call would be acceptable rather than always having to specify times individually in five-minute intervals.

Extra Facilities. It was necessary for a wide number of other facilities to be available on the system, which would be selected by entering codes into the command input ORD II field. (e.g. supervision of coinbox calls). The inquiry field 'IN' was to be designed to enable the operator to call on a number of specialists to help with problems, to contact a supervisor, etc.

The miscellaneous field MISC was considered to be useful for entering additional text notes about a call, (e.g. so that operators subsequently dealing with it would understand the special circumstances involved).

B.4.2. Format of Datascreen

The layout of information on the call tickets was considered to be an important factor, likely to affect the performance of operators markedly. The datascreen ticket was to provide information to the operator, on the nature of the call, the state of subscribers and the history of the call. Consequently, it was considered that the relevant display fields within this ticket should be structured to aid search, to clarify the call status, and to aid decision-making. The ticket was also intended to provide a channel of information from the operator to the system through unprotected fields which were to be completed by using the keyboards. Access to the required field, by identifying it with the cursor, had to be easy, and editing functions had to be simple. As text was entered it had to immediately appear, delays of greater than 0.5 s being considered unacceptable. Text items which had yet to be verified by the system and those unacceptable were recommended to be distinguishable from accepted text strings; appearing in a brightened form when first entered and with normal brightness once accepted.

The ticket which had been originally designed for ANE 403 by LME personnel (see Fig. 8) had several disadvantages. Firstly, the fields had been organised into groups by size rather than function. The main fields had been organised in the sequential order in which they were completed, except that the fields indicating the status of the A and B subscribers appeared beneath the fields showing the respective telephone numbers. The short fields appeared to follow no clear pattern, with time information (TOR - time of order, TOS - time of start, TIM - length of call) scattered among operator signatures, (SBO, SCO), price advice (PA) and call number (CNR). A large number of lines had been reserved for miscellaneous text input (MIS), so restricting the space available for more important fields. The fields tc be completed by the operator (indicated in Fig. 8 by dotted lines) had been scattered among protected fields, making cursor movements to access them more difficult, and making it necessary for the cursor to appear in protected fields (to simplify the software and prevent 'jumping'), even though text could not be entered into these fields.

The redesigned ticket recommended by the author shown in Fig. 9, started with a field indicating the nature of the call (e.g. DEMAND, DELAY, etc.). The ORD I field, with instructions from the computer to the operator, was also placed on the first line, so that, when a ticket appeared, it would immediately be obvious what type of call it was and what the operator was expected to do. The next two lines of the left hand side were devoted to the type of call (TYC) and the categories of priority it had (CAT). The fourth line (REM) provided space for the computer to include remarks, such as the number of times a call had been attempted (e.g. 3RD ATTEMPT). On the right had side of the screen were four protected fields containing information relevant to call handling. The field to the right of line 1 was for the system to enter the cost of a call when price advice had been requested (PA). This field had to be longer than the fields on the next three lines and therefore its label was shifted left to accommodate it. The field to the right of line 2 would record the length of a call (TIM) and was to be read by the subscriber in conjunction with price advice. The "call number" (CNR) field on line 3 was for the entry of a number to identify the call uniquely. This number could be entered automatically by the system or by the operator via the command input ORD II field. RET was a field reserved for entering the "recall time" of a delay call, so that, when a system alerted an operator to deal with the call, he could note this time.

On the redesigned ticket all the fields open to the operator were placed within a section in the centre of the screen uninterrupted by protected fields so that cursor movements could be limited to these unprotected fields. All the unprotected fields, except one, began at the left hand side of the screen, so minimising complex cursor movements, since in most cases "line-feed-up" or "line-feed-down" would access any field. The inquiry field (IN) was provided for the operator to enter the codes of relevant enquiry services which would then be contacted by pressing the inquiry button (IN) on the special function keyboard. The lack of screen space made it necessary to move one field across to the right hand side of the screen, so the least used field "inquiry" had to be placed at the end of the ORD II field which could be made relatively short. Although it was likely that the ORD II field would not be used on all demand calls it was possible that it would frequently be needed on delay calls and therefore it was placed at the top of the unprotected fields.

The calling number (A-NR), called number (B-NR) and called person (PER) fields were longer on the redesigned ticket, providing more space for names, extension numbers etc., and enabling the operator to space numbers for easy reading. The miscellaneous field (MIS) was much shorter on the new ticket. Only two dotted lines appeared for typing but the dashed dividing line could also be over-written, providing a total of three lines of 64 characters, which was considered to be sufficient. At the bottom left of the ticket a graphic indication was provided showing the state of the calling and called subscribers connection to the operator. In the centre, the status of the individual subscribers as detectable by the system appeared in word form (e.g. BUSY, NO ANSWER etc.). The time a call was ordered (TOR), its start time (TOS), and the signatures of the booking and completing operators (SBO and SCO respectively) appeared to the right hand side, since they would be infrequently needed. The final two lines of the ticket (labelled INF) were for messages from the chief supervisor to individual or group operators (e.g. informing them that certain lines are busy). It was considered that it might be beneficial if the chief operator were able to select 'reverse video' for important messages, but it was also stated that the supervisors should be encouraged to use such a facility sparingly to avoid unnecessarily distracting operators.

### B.4.3 Keyboard Design

It had been determined that the keyboard which was to be used with ANE 403 would consist of four separate groups of keys (see Section A.6.). This section of the literature review dealt with the layout of keys within each group, and with the layout of the four groups with respect to each other.

It was stressed that the position of keyboards and the layout of keys within them was important, and was likely to affect performance markedly. It was recommended that the keys with related functions be placed in groups if this proved possible without conflicting with other task demands (e.g. sequential operations). Keys which had far reaching system effects were required to be carefully placed and to be spaced from other keys to reduce the possibility of their accidental operation. It was recommended that keys conform to operators' expectations (stereotypes) with regard to position and operating method. The keyboard design which had been prepared by LME engineers is shown in Figure 10 and, as described below, a redesign was suggested based upon ergonomics principles. The redesigned keyboard is shown in Figure 11. The circled letters in this figure indicate the following groups of keys.

A. Alphabetic Keyboard. The alphabetic keyboard of the original design was arranged in the standard QWERTY arrangement as on typewriter keyboards (see Fig. 10). Despite the design of several alternatives, this keyboard was by then almost universally used in countries using the A-Z alphabet and had been accepted as the International Standard. Several additional characters were needed in some countries (e.g. Å, Ä, Ö in Sweden) but again standard layouts usually existed. Keyboards laid out in alphabetic order (i.e. A, B, C ... Z) had not been found of any distinct advantage especially with experienced users. The QWERTY layout was therefore recommended for the redesigned keyboard (see Fig. 11). The additional characters such as punctuation marks, currency signs etc. were also recommended to follow common layouts. The numerical keys were eliminated from the new design, however, since a separate numerical keyboard existed. It was considered feasible for the shift function to be avoided in the new design, and in any case the shift lock facility found on standard typewriters was not recommended for ANE 403 because this was considered likely to increase the likelihood of errors.

B. Numeric Keyboards. At the time of the report a number of standard numerical keyboards existed, the most common being the ten key 3x3+1 layout. Unfortunately, this keyboard was widely used in two distinct 'standard' forms; the 'ADD' keyboard which was used on calculators and adding machines had 7, 8 and 9 on the top row, while the touch telephone 'TEL' keyboard had 1, 2 and 3 on the top row (see below).

"ADD"

| | | |
|---|---|---|
| 7 | 8 | 9 |
| 4 | 5 | 6 |
| 1 | 2 | 3 |
| | 0 | |

"TEL"

| | | |
|---|---|---|
| 1 | 2 | 3 |
| 4 | 5 | 6 |
| 7 | 8 | 9 |
| | 0 | |

The differences in performance on these two keyboards were minimal when comparisons were made between users experienced in using either one or the other, although population stereotypes favoured the TEL keyboard. It was expected that ANE 403 operators would use only one type of keyboard and would not be expected to operate calculators as a normal part of their job. In addition, telephone operators were anticipated to have become skilled on the TEL keyboard which was used on systems current at that time. The use of the TEL keyboard had been incorporated in the original design (Fig. 10) and was recommended in the new design (Fig. 11).

However, on the original design no handrest was incorporated below the numeric keyboard. It was considered probable that the numeric keyboard would be operated with a great degree of skill, often without looking at it, and it was therefore recommended that a handrest be provided, so that the heel of the hand could be steadied while keying took place. It was required that this handrest be at least 90mm x 60mm. Such a handrest is shown in the keyboard design shown in Figure 11.

C. Editing Keys. The editing keyboard originally suggested was provided to enable text editing to be carried out (see Fig. 10). The editing functions are dealt with separately in Section B.4.9 since they were to require datascreen fields as well as keyboards.

D. Cursor Control Keys. Two functions were to be conducted using the cursor control keys; it was to be possible for the cursor to be taken to the beginning of the relevant fields for the entry of data or of orders to the system (field selection), and the cursor was also to be used to select a particular point in the text for subsequent editing.

Field selection could have been designed to be carried out by any of a number of conventional means:-

1. Standard cursor movement keys only - up, down, left, right, etc.
2. Standard keys with the addition of special cursor movement keys such as 'home', new lines, and tabulation (this was the solution adopted in the original keyboard design shown in Fig. 10).

3. The provision of buttons marked with the names of the fields which, when pressed, would immediately take the cursor to the relevant field.

4. Entry of simple field identity codes before data entry which would cause the cursor to skip to the relevant fields.

5. Automatic cursor movements to place the cursor in relevant fields when certain procedures were being carried out or certain orders were given by the system. (This suggestion is dealt with further in Section B.4.8 since it does not involve the keyboard).

However, since it had been possible to organise all but one of the operator accessible fields such that they all began at the extreme left of the screen (Fig. 9) any field could be accessed by cursor moves similar to 'carriage return line feed' but moving either up or down. It was therefore recommended that the vertical movements of the cursor were to be controlled from the top and bottom keys (↰, ↴), which were also to cause the cursor to move to the beginning of the field so accomplishing the desirable 'carriage return' function as well as a 'line feed' up or down. It was decided that these keys would not need to be 'typomatic', since the redesigned ticket only required a few depressions of these keys to reach any unprotected field (see Fig. 11).

The 'home' key (↖) in the centre of the four other cursor keys was retained from the original design (Fig. 10) to enable the operator to select the ORD II field simply and easily since it was to be into this field that the operator would enter 'commands' or 'orders' to the system.

Once within a field, the cursor was to be used to indicate characters or spaces so that erasures or inserts could be made. In the original design, shown in Figure 10, a number of cursor controls were suggested, but it was suggested that for this purpose, a simple left or right cursor control was sufficient, providing that its speed would enable fast access of the required position (see Fig. 11). In applications existing at the time of the review it was known that the users could access particular places in text by specifying a string of search letters and the system would immediately search for the occurrence of that string in the text and place the cursor in the appropriate position. For the ANE 403 system the text lines were recommended to be short and it was considered unlikely that a system of the type described above would prove beneficial.

It was realised that the character selection procedure system had to be reccmmended largely based upon the dialogues envisaged and upon the format of information on the screen. It was recommended that it the screen layout described in Section B.4.2 was used, a simple 5 function cursor movement keyboard would be sufficient (see Fig. 11).

In the keyboard design shown in Figure 11, it was recommended that the cursor keys at the left and right moved the cursor one space at a time in the relevant direction if momentarily pushed. It was also suggested that these keys were 'typomatic', that is the movement of the cursor would become continuous if the key was held down for longer that 0.5 s. The movement of the cursor under the 'typomatic' mode was recommended to be fairly slow in order to enable accurate positioning of the cursor for editing (e.g. 10-15 characters/s). The cursor was recommended to halt at the beginning or the end of datascreen fields when the typomatic function was used. For example, on line 1 of the screen format it was suggested that the cursor should only move to the end of the operator order or command (ORD II) field until the key was re-pressed to make the cursor jump to the beginning of the next field (i.e. the Inquiry 'IN' field).

E. Special Function Keyboard. The special function keyboard was necessary in order to enter instructions into the computer. In the original keyboard (Fig. 10), some functions were to command the computer to carry out exchange operations e.g. Split to calling subscriber (SPA), initiate call to calling subscriber (A), clear line to calling subscriber (CLA). Another function "TRANS" was to order text to be transmitted from the terminal controller to the central computer where it would be interpreted. The function button (IN) was intended to initiate contact with an inquiry system within the exchange or with a supervisor. Two buttons were necessary to enable call time to be controlled, "TS" (for "Time Start") was to start or stop the clock, while "RESET" was to reset it to zero. The final button "FREE" would inform the system that the call handling had been completed by the operator. The "ABC" key, (ABC was an abbreviation for Auditory Broadcast call) was included to enable it to be used in certain installations. When this key was to be operated a new call was to be indicated to the operator by loudspeaker and not through the headset. In redesigning the special function keyboard it was stressed that it was important for the layout of the function keys to follow stereotypes or common expectations. The layout which was finally suggested is shown in Fig. 11. On the left are shown the seven exchange control buttons, with those controlling calls to the calling (A) subscriber on the far left of the kwyboard (A, CLA, and SPA) and those controlling calls to the called (B) subscriber on their right (B, CLB and SPB). The "speech monitor" button (SM) was made larger than the others and was placed between and at the foot of the rows of keys for the 'A' and 'B' subscribers. This layout not only corresponded with the standard practice and an intuitive stereotype but, also with the symbolic representation of call state which had been suggested for inclusion on the screen (see Section B.4.2). To the right hand side of the exchange control keys and other function keys were positioned according to their relative frequencies of use and priorities. The "RESET TIME", "inquiry" (IN) key and "ABC" (Auditory Broadcast loudspeaker) keys were unlikely to be used frequently and were therefore placed at the top of the keyboard,furthest away from the operator. The "TRANSMIT" and "FREE" keys were made larger than the others and placed one above the other at the bottom. The "TIME START" button was likely to be operated immediately before freeing the terminal and consequently was placed in the centre of the keyboard above "TRANSMIT" and "FREE".

B.4.4. Position of Keyboard Groups

The four groups of keys had to be positioned carefully with respect to each other and to the operator. The original design suggested by LME designers is shown in Fig. 10 and the design recommended by the author is shown in Fig. 11.

The size and importance of the alphabetic keyboard made it necessary for it to be mounted in the centre of the console where it would be useable by both hands. The alphanumeric keyboard was quite deep and to reduce substantial arm movements it was recommended that this be brought near to the front edge of the desk. The numeric keyboard was considered equally important on this system but it was clear that its small overall size would enable it to be operated most easily with the heel of the hand resting below it. Consequently it was mounted higher up the keyboard with some space below it. On the design recommended, the numeric keyboard was placed on the right hand side for two reasons.

1. To produce a more balanced design.
2. Because people's right hands are usually stronger, more accurate and faster than their left hands.

The cursor movement keys could have been placed to either side of the operator but in the design shown they were placed to the right to create a more even distribution of keys around the alphabetic keyboard. Again a hand-heel rest was considered to be beneficial for improving "blind" operation of these keys (e.g. when the operator viewed the screen).

The function keys could have been placed on either side of the keyboard from the point of view of skilled operation, since the keys were to be operated singly and were well spaced. The screen format already suggested had a section related to the status of the exchange on the left hand side, (see Section B.4.2) and it was considered preferable for the keys relating to this information to also be placed on this side. Since these keys were to be placed on the left hand, it was recommended that more care was taken when designing the spacing between them, because the accuracy of left hand keying was known to be usually worse than that of right hand keying.

The keyboard layout which was recommended is shown in Fig. 11.

## B.4.5. Labels and Abbreviations

The limited screen area, and the small number of characters which it was possible to display, made it necessary to abbreviate field labels, instructions and place names. However, it was anticipated that, on a system operated by experienced personnel, misunderstandings of frequently used order codes would be very rare and that the fields would be known from their positions on the screen rather than from the label. Care still had to be taken, however, to ensure that the abbreviations used were unambiguous and straightforward. Several authors had suggested rules for the generation of fairly acceptable English abbreviations, but these could only be used as the starting point in any particular installation, (Bourne and Ford, 1960). One point to which careful note was made, was that it was preferable for abbreviations to make it clear whether the abbreviation was devised from the initial letters of several words or whether it was merely the first few letters of a single word. The abbreviations previously suggested by LME engineers had been inconsistently derived and were ambiguous.

| Code suggested before study | Meaning | Code recommended by present author |
|---|---|---|
| TIM | Time | DURATN (Duration) |
| RET | Recall Time | RECALL |
| TOS | Time of Start | BEGIN |
| SBO | Signature of Booking Operator | SIGN 1 |

* Underlinings indicate the derivation of the previously suggested code.

It was recommended that where possible the abbreviations used in each country were to be devised by a native of that country who would understand the job, would have spoken to existing operators and would understand some ergonomics. A complete list of the recommended English codes for ANE 403 is given in Appendix 1.

## B.4.6 Coding Methods

A number of techniques could have been used to highlight, and thereby to draw attention to, specific pieces of information. Blinking or flashing characters of cursors were considered extremely useful to attract attention but it was cautioned that they be used wisely. It was pointed out that, in practice, only a very few blinking characters could be dealt with at any time and that a maximum of three blink rates per system should not be exceeded. In the current system, ANE 403, it was proposed to blink the cursor at 3Hz to ensure that it could be found on the screen easily (see Chapter B.2.4.). In view of the suggestion to use a blinking cursor no other blinking codes were to be used.

It was recommended that selected characters could be displayed brighter than others to make them conspicuous. Like blink coding, a maximum of three coding levels was recommended and it was also stated that they should lie on a logarithmic rather than a linear progression. For ANE 403 LME engineers had proposed that newly entered characters, not yet transmitted to the central computer, would appear brightened until accepted. The present author considered that the technique was quite useful, but pointed out that the brighter image had to be significantly more intense than other characters, At the time of the literature review it was considered possible that a large enough intensity level might not be achievable, given the relatively high base-level required, but subsequently it was found possible to achieve the required brightness difference with an adequately high base-level. It was not considered necessary for field labels to appear at a third brightness level. Brightness levels were recommended to be limited to two which were to be in the ratio of 4:5 (e.g. 120 and 150 $cd/m^2$).

As described in a previous section (see B.2.5.) it was recommended that light characters on a dark background be used for the majority of the displayed text. Reversing the direction of contrast of certain fields or parts of fields would produce 'black on white' characters which were known as "reverse video". Although the technique was likely to prove extremely useful in some applications, especially for attracting the operator's attention, it was felt that the high brightness of such areas might have distracted ANE 403 operators from other information. The only field when "reverse video" was recommended was in the "information field" where important information was to appear, relevant to operator's work and specific to an individual or pre-specified group of operators. In this case the "attention getting" capabilities of reverse video were to be exploited.

### B.4.7 Lock-outs and Protected Fields

It was known that one way to prevent operator errors from having serious deleterious effects on systems was to prevent them from acting in dangerous ways. Aircraft controls for instance were frequently inter-locked to prevent actions out of sequence. In the present application, it was necessary for certain "data fields" to be protected from input or editing by the operator, while "field labels" had to be protected both from the operator and from the system. The operator was also to be prevented from "freeing" a ticket until all the required information has been entered ( so that, for example, a delay call could only be booked if a time had been entered into the RET field via the ORD II field). It was recommended that interlocks and protected fields be provided for ANE 403.

B.4.8 Automatic Cursor Movements

In a great many cases, it would have been possible for the cursor to be automatically positioned in the next relevant data field without intervention by the operator. For instance, if the system had asked for the signature of the operator when the terminal was opened, the cursor could have been pre-positioned in the relevant ORD II field by the system. This technique might have been used to advantage, but on the new ticket design a high implementation cost could not be justified. It was thought that a short experiment could have quantified the time which might have been saved by automatic cursor movement. However, in the event, this experiment was not conducted.

B.4.9 Editing Functions

The editing keys were considered necessary if the operator was to be capable of correcting text already entered onto the screen. Although it would have been simpler to require that the operator erased complete fields and then re-wrote them, it was possible that in retyping a field other mistakes might have been made. More comprehensive editing functions were recommended.

Of the editing functions included on the original keyboard (see Fig. 10) "insert line", "delete line" and "erase line" were not considered necessary and it was stated that these could be omitted from this design.

Erase Field. This function was to enable the operator to delete the whole of an unprotected datascreen field. It was recommended that the "erase field" function should not depend on the position of the cursor within the field. This function was recommended for ANE 403.

Delete Character. This function was to remove the character under which the cursor was positioned and "to close up" the surrounding text. It was anticipated that this function would be most useful in the correction of words and, in particular, names but would not be needed for telephone numbers, which it was thought would seldom need to be 'closed up'. The cursor was required to remain in the same screen position after character deletion (e.g. under the next character in the string). This function was recommended for ANE 403.

Erase Character. Although similar to the "delete character" function, the "erase character" function would have removed the character above the cursor without 'closing up' the text. Strictly it was not considered necessary since the operator could simply over-write the character with a space. If long strings of characters had been needed to be removed it could have been made typomatic (i.e. it would erase strings of characters if held down for any time). This function was NOT recommended for ANE 403.

Erase Left. It was known that during data entry it is common for people to detect errors as they make them and therefore it was recommended that facilities should be provided for these errors to be corrected immediately. The 'ERASE LEFT' key was considered to be extremely useful in such situations, allowing the operator to progressively erase numbers or letters to the left of the current cursor position. This function was recommended for ANE 403.

Insert Character. In the same way that it was anticipated that operators may have to delete an existing character and close up the text it was also likely that he may occasionally wish to insert a new character into an existing string. The insert character facility required a 2-stage operation, selection of the function and then entry of a character.

It was recommended that the depression of the "character insert" key should cause a character space to be inserted above the cursor by moving the text to the right by one position. Since the space was effectively inserted to the left of the character under which the cursor was originally positioned it was suggested that the cursor should have a short vertical line at its left hand end to make this clear.

EXAMPLE

| | |
|---|---|
| Starting position | STCKHOLM_ |
| Cursor re-positioned | STCKHOLM |
| Character insert key depressed | ST_CKHOLM |
| 'O' key pressed | STOCKHOLM |
| Cursor re-positioned | STOCKHOLM_ |

Recommended for ANE 403.

Replace Character. No special key was needed to replace one character by another. If a character key was pressed it would replace any other character under which the cursor was positioned. No special 'replace character' function key was recommended for ANE 403.

B.4.10. Operators handling partial calls

A great many international telephone calls required later intervention by the operator; to handle delay calls, to give price advice etc. On ANE 403 the ticket accompanying the calls was originally designed to be potentially available to any operating position and the system would allocate these later transactions to any free operator with the necessary skills (e.g. language spoken). In such a design it was considered to be unlikely that operators initiating calls would deal with subsequent call handling. While this system would have provided a great deal more flexibility than was present in more traditional exchanges, and may have resulted in subscribers being dealt with more quickly, it was considered to carry potential penalties. In any system it is frequently found that unusual circumstances arise which have not been predicted by the designer and with which the system cannot cope without human intervention. On ANE 403 it was considered that it would be beneficial if operators could label certain calls in some way such that these calls return to them wherever possible for subsequent call handling operations to be performed.

## B.5. CONSOLE DESIGN

It was necessary for the datascreen and its associated keyboard to be mounted on some form of table or console such that these devices could be used easily.

### B.5.1 Location of Screen

The screen size, its position relative to the operator and its angle of presentation had to be decided with respect to a number of factors; legibility, flicker perception, eyesight problems and ambient illumination. The first three of these topics are dealt with in this section, ambient lighting is considered in Section B.6.

Legibility. It was apparent that generally the legibility of characters from a given set would be determined by the visual angle they subtended at the observers' eye. The relationship between visual angle and legibility was not considered linear, however and, although larger characters were considered to be individually more legible, beyond a certain point greater scanning angles were thought to require greater eye and head movements in order to read whole words and consequently, therefore, legibility of text was likely to decrease. As described in Section B.2.1, character heights subtending 18-22 minutes of visual arc were chosen as reasonable for the task. The sizes of screen thought to be required to provide adequately sized characters at three viewing distances are described in Section B.2.7.

Flicker. The degree of flicker which it was considered would be perceived by a given population of users with the equipment provided was thought to be largely determined inter alia by the total brightness of the screen and the solid angle which would be subtended by the screen at the eye of the operator. The screen sizes which had been suggested for various viewing distances subtended similar solid visual angles and therefore would show no differences in flicker. Further observations made on the problems of flicker in relation to light fittings, etc. are dealt with in Section B.6.4.

Eyesight. No attempt was made in this phase of the work to describe eyesight problems fully, since they were covered in Bedwell and Stone (1976), a document which accompanied the original literature review. The main problems which would affect console design were thought to be those related to acuity, accommodation, convergence and visual scanning.

Many experiments described in the literature reviewed had related character size to 'good' or 'well-corrected', vision (so called 6:6). The character heights subtending 18-22 minutes of arc were considered to be adequate for up to 6:24 eyesight.

Accommodation is the name given to the ability of the eye to alter the focal length of its lens, such that light from objects at different distances from the eye can be focussed on the retina. Accommodation is accomplished by changing the shape of the eye's elastic lens by means of muscles around its periphery. As the eye ages, the lens becomes less elastic and consequently progressively less able to focus on near objects without suitable optical corrections.

It was essential that the range of eye heights found in the seated population be accommodated in the console design for ANE 403. Three factors were considered; probable distance of eyes from the screen, angle of declination of eyes when reading the screen, and angular movement of the direction of gaze when moving the gaze from the screen to the keyboard. The distance of operators from the screen has already been covered (Section B.2.7) and may be summarized by saying that it was recommended that characters subtend 18-22 minutes of arc at the eye (provided accommodation difficulties were dealt with). The angle of declination of the eyes to the screen was considered important for two reasons. Firstly, it was essential to maintain comfort, given normal head slumps and eye movements. Secondly, it had to be easy to view the screen through the lower half of bifocals (if worn) without uncomfortable head movements. In general, primary displays (e.g. datascreen) were recommended to be kept within an arc $30^{o}$ below the horizontal through the operators eyes (this assumed a $30^{o}$ solid angle at the eyes about a normal line of sight $15^{o}$ down from the horizontal). It was also recommended that secondary display (e.g. keyboard) were to be kept within an arc $45^{o}$ below the horizontal through the operator's eyes. At most it was essential that the angular movement at the operator's eyes, when glancing from the datascreen to the bottom of the keyboard, was kept within $45^{o}$, allowing $30^{o}$ for the comfortable range of eye movement with a fixed head position and $15^{o}$ of head movement.

Recommendation. It was recommended that the base of a 10" datascreen (measured diagonally) be placed 180mm above an adjustable height desk and 300mm from its front edge. The screen was to be fixed at an angle of 22° from the vertical such that it was approximately perpendicular to the angle of sight. If adjustable the screen angle adjustment was recommended to range from 0-30° to the vertical.

B.5.2 Keyboard Position

Distance forward on desk. The keyboard had to be enclosed in its own case, to provide the necessary angle of presentation (see Sections B.3.6 and B.3.7) while enabling it to slide fore and aft sufficiently to account for anthropometric differences in forearm length. It was recommended that the keyboard case should not be capable of lateral movement on the desk because it was considered important to encourage operators to sit square to the keyboard and on the median line of the screen.

Height and Angle. The height of the keyboard was recommended to be such that the forearms would remain horizontal or would slope slightly down when keying. As the hands moved to the upper part of the keyboard it was necessary for the slope of the keyboard to follow the upward movement of the hands while the forearms remained approximately horizontal.

Thickness of keyboard and table. To provide sufficient leg room under the surface supporting the keyboard at the correct height, the front edge of the keyboard and the table was recommended to be as thin as possible. For this reason the thinner of the two Rafi keyboards was recommended in the earlier Section, B.3.6.

It was thought that the keyboard would act as a secondary display when the operator looked at the keys. Consequently the keyboard was recommended to be kept within a viewing angle less than $60^{o}$ vertically down from a horizontal plane through the operator's eyes. In the horizontal plane the edges of the keyboard had also to be kept within an angle of $\pm 30^{o}$ from a vertical plane drawn through the centre of the operator's head and the centre of the screen.

Recommendations. The keyboard was to be enclosed in a case such that it could slide fore and aft but not laterally. This mechanism had to enable the 200mm deep keyboard to be moved by at least 100mm fore and aft. In its foremost position, the front edge of the keyboard was to be able to reach the front edge of the desk. In its rear-most position, its back edge would ideally be directly beneath the base of the screen.

### B.5.3 Chair Design

One of the most important elements thought to determine the adequacy of the console was the chair. This not only would support the operator's weight, but was to be critical in the placement of hands and eyes in relation to the task, and thus cruicially determining posture. The ideal chair for ANE 403 was thought to be a typist's chair, following ergonomics standards of size, shape and range of adjustment.

It was suggested that the chair should not incorporate any armrests because their benefits were thought to be limited, because their adjustment was usually difficult and because it was possible that they would hit the desk.

Two ergonomic recommendations the available for seat design are shown in Figs. 12A and 12B. Figure 12A shows the minimum specification existing for the dimensions of office chairs as defined by the British Standards Institution (1965). Figure 12B gives the recommended values for seat design which had been specified by Diffrient et al.(1974).

B.5.4 Adjustability of Console and Chair.

Why adjustment? It was considered the ideal to provide sufficient range of adjustment for all items of equipment to accommodate at least 95% of the population expected to use them.

Once and for all. In a situation where only one individual would use a workplace, it would have been considered necessary to provide an initial degree of adjustment to adapt the equipment to the individual. However, in such a situation, once the adjustment had been made, it was possible that it would never again be used and the adjustment mechanisms would be largely redundant. In order to reduce the costs of necessary, but infrequently used, adjustment mechanisms it was recommended that, for such situations, they be designed for cheapness rather than for ease of use. However, this was not the type of use envisaged for ANE 403.

Frequent adjustment. In a situation, such as was proposed for ANE 403, where several people would share the use of a single piece of equipment at different times of day, adjustment was considered to be problematical. To make equipment easily adjustable, not only had each adjustment to be easy to make, but the number of such adjustments had to be small. If adjustment of one dimensions was made possible this would affect the need for, and the range of, adjustments in other dimensions.

It was considered that the operators would have to be shown how to adjust the equipment; and the importance of each dimensions would have to be demonstrated to them. It was advised that it would be beneficial if, having found an acceptable console arrangement, operators were able to duplicate this position at any console by reference to rules or marks on the console itself. It was stressed that in the initial periods of operator training it would be essential for operators to be trained to use the adjustments effectively.

It was further recommended that periodically, supervisors were to be encouraged to check that adjustments were being made and that bad postures were not being developed.

Recommendation. Providing that adjustment mechanisms could be designed which were fast and easy to use, it was recommended that they be included. It was recommended that training programmes be instigated to help operators learn to adjust their chairs and consoles to the correct dimensions. The adjustment ranges of desks and chairs recommended were:-

| | |
|---|---|
| Chair seat height: | 360-520 mm (minimum 380-480 mm) |
| Desk height: | 600-720 i.e. 120 mm adjustment |
| | 560-760 including leg adjustments for very extreme individuals and for wide variations in populations (e.g. Japanese) |

B.5.5 Storage

It was recommended that space be provided for storage of directories, etc., necessary for the operation of the system and that sufficient workspace be given for these items to be used easily.

The console designer was reminded that operators would bring with them minor items necessary for the job (pens, etc.) and various personal items. While it was suggested that some would be carried in pockets or handbags it was pointed out that the operators might wish to leave them in personal lockers from day-to-day. Personal lockable drawers in each console were considered desirable but, in the event of this being impractical, lockers were suggested for work areas or the rest rooms.

Recommendation. It was recommended that the storage of documents, writing instruments and personal items be considered in the design of the workplace and/or the building.

B.5.6 Visual Panels

It was pointed out that tall panels between individual operator terminals would reduce visual and auditory distraction but they were not recommended in situations such as would exist on ANE 403, where operators were to be required to sit at their terminals for long periods. To provide visual variety and to enable verbal communication between adjacent operators, large inter-terminal panels were to be avoided. Higher panels between working groups or around rest rooms were recommended, however, to provide a more 'friendly' environment and to reduce noise levels.

B.5.7 A Suggested Console Design

While the recommendations made thus far were put forward to be considered by console designers they did not themselves generate a unique design solution. In this section a console design was suggested which appeared to satisfy the majority of the human factors requirements covered in the whole report. At this stage however, it was pointed out that this design could not be put forward as a definite recommendation until fully evaluated and modified as necessary.

The design was quite simple, consisting of an adjustable height desk supporting a tilting screen and a sliding keyboard (Figs.13A-C). The design was to be capable of accommodating users whose body dimensions would differ markedly. Figure 13A includes the relative positions of important body parts, and the values of the links between them, for a 2.5 percentile US female, and a 97.5 percentile male (data - Diffrient et al. 1974). Figure 13A was drawn with the users hands and the keyboard taken as fixed points of reference for the other parts of the body and the other parts of the console respectively. If the operator were to adjust the chair for general comfort, and then adjust the desk height to bring it just below elbow height, the screen fixed to the adjustable desk appeared to be satisfactory. The small differences in viewing angle for the population extremes under these conditions suggested that a fixed screen angle might prove acceptable, but that some adjustment should be provided to enable unwanted reflections to be removed.

It was suggested that the keyboard could be mounted between two parallel 'rails', enabling it to move back to the datascreen or forward to the edge of the desk. Again, although an adjustable position keyboard did not seem necessary from the anthopometric data, individual preferences for keyboard position, and the postural reliefs given by moving the keyboard occasionally, were considered sufficiently important to warrant this flexibility.

In elevation, the screen did not lie entirely within the angles of view recommended in B.5.1 and B.5.2, but were considered sufficiently close to them to prove acceptable for at least 95% of the population. In plan (Fig. 13C), the screen and the keyboard subtended angles of 24$^{\circ}$ and 53$^{\circ}$ respectively at the eyes of the 2.5th and 97.5th percentile of users, at, or a little over, 500mm from each of them. These were, however, within the outer limits of 30$^{\circ}$ and 60$^{\circ}$ specified for primary displays (Section B.5.1) and secondary displays (Section B.5.2.).

It was recommended that the desk had to be at least 610mm wide to provide sufficient kneehole space (Diffrient et al, 1974) but, finally a 1,000 mm wide desk was suggested, to enable the operator to enter and leave the console more easily. The longer desk also allowed provision for storage on the desk itself, where it would be easily reached.

## B.6 LIGHTING

The visual tasks which were to be involved in operating ANE 403, and the predicted difficulties of maintaining high contrast on the datascreen under normal lighting conditions, made it imperative that particular attention was paid to the design of room lighting. The position of light fittings, the subsequent direction of light and the distribution of light within the field of view were all considered as important as light intensity. These factors are dealt with in this sections, together with a description of the causes of flicker and means of reducing it.

### B.6.1 Light Level of Intensity

Dunn (1972) has examined the lighting of radar rooms, recommending a level of only 50 lux, but Stocker (1964, 1966) had suggested that, if printed materials needed to be consulted occasionally, 150 lux was more appropriate. The Illuminating Engineering Society Code (IES, 1973) had recommended a level of 300 lux for cordless telephone switchboard rooms. Hultgren and Knave (1974) in their survey reported that a level of 300 lux had caused no visual problems for the operators of well placed datascreens. The problems at other than well placed datascreens were explained by directional effects from natural and artificial lights. It was recommended that, for ANE 403,300 lux of illumination would prove sufficient for all terminal tasks without the need for higher levels of local lighting. It was noted, however, that, if lower lighting levels (e.g. 200 lux) were considered necessary in any particular situation, then this should be supplemented by local light sources (of up to 500 lux) for reading printed materials. Stress was placed upon the importance of assuring that such local lighting was designed to be glare free. It should be noted, however, that the recommendations made above were based entirely upon the literature review. Subsequently, experiments conducted in Phase II of the project demonstrated that the datascreen and console design could be used in a light level of 500 lux (see Chapter C).

The absolute light level striking and being reflected from, the working plane could not be considered alone. Hopkinson and Collins (1970) stated "it has been shown that better lighting results when the whole pattern of the visual field is studied". Therefore the lighting was not only to be adequate for the task but to be pleasing whenever operators looked away from the datascreens. The I.E.S. Code (1973) recommended that the ratio of the minimum to the average illumination levels over working areas with general lighting should normally not exceed 0.8. The I.E.S. Code also recommended the desired distribution of the lighting on the bounding surfaces of rooms, since this was known to contribute importantly to the overall appearance. Low uniform light levels would have reduced the masking effects of reflections in the datascreens, but might have made the room dull or even soporific. It was advised that interest be introduced into the interior design of operator rooms by the careful use of colour contrast.

For ANE 403, a uniform light level of 300 lux was recommended, but it was stressed that, in specific installations, careful consideration had to be given to lighting design in relation to console positions. Gross contrasts within the field of view were to be avoided. An alternative lighting scheme was suggested, involving general lighting levels of about 200 lux with glare-free local sources of up to 500 lux.

B.6.2 Reflectivity of Surfaces

The visual environment was known to be affected by the reflectivity of surfaces and the I.E.S. Code (1973) had suggested the reflectivity most desirable for walls, ceiling and floor; as well as for working surfaces. It was thought that these reflectivities should provide a reasonable visual appearance to surroundings, without deep shadow and harsh contrasts.

Within the workplace area it was recommended that the surface around the task should have a reflectivity of only 1/2 to 1/3 that of the task areas (IES, 1973). For the ANE 403 system, however, it was known that a major portion of the task would consist of viewing phosphorescent characters on a dark screen. If the screen background was to be made as non-reflective as possible, to enhance contrast, it was considered possible that it would prove difficult, in practice, to make the screen surround 1/3 as reflective. It was recommended that the screen surround be coated in low reflectance material, preferably with a matt randomly textured surface.

B.6.3 Glare

It was explained that glare could be experienced as discomfort, or it could cause a loss of performance, but that it might be largely eliminated by careful design and siting of equipment.

In using the datascreen it was possible that the operator would be subjected to two primary sources of glare, direct from light fittings or indirect glare from specular reflections from the datascreen. The latter were considered likely to be particularly disruptive because they might mask information on the datascreen.

While windows were thought to be attractive, and it was likely that they would be liked by operators, there was a major risk that they would prove to be a prime source of glare. Although a windowless room would enable total light control to be exercised, it was stressed that care had to be taken to compensate for lack of windows by creating an interesting visual environment. It must be noted that this recommendation was rather overcautious because it was to be passed on by L.M.E., the manufacturer, to the purchasing telecommunications administration who could then pass it to the architect and building contractor responsible for the operator room. It was felt that it would be difficult to pass complex, interrelated conditional recommendations down through this chain.

It was recommended that the Glare Index from light fittings as defined by the IES Code (1973) did not exceed 16 (Östberg, 1974).

B.6.4 Flicker

All forms of alternating-current lighting were known to give rise to frequency variations in intensity and colour, some of which could be perceived as flicker. Large number of people had reported that they felt that artificial light was worse for the eyes than daylight (Wells, 1965). Flicker was often considered the major source of dissatisfaction with artificial light (Brundrett, 1974).

In general, the flicker fusion threshold, the boundary between an observer seeing or not seeing flicker, was reported to be affected by a variety of factors; frequency, waveform, target luminance, and, for small targets, background luminance. The frequency at which a light source fed by a square wave supply fuses, and appears to be continuously 'on', is termed the critical flicker frequency (CFF). The CFF varies logarithmically with the luminance of the target and with the area of the target on the retina. For small fields ($< 1^{o}$ diameter) the fovea was known to be most sensitive but it was also known that large fields were more likely to give rise to flicker if seen peripherally.

It was known that the frequency below which flicker was perceived varied widely in the population, and within individuals it was known to vary due to physiological and psychological factors. The degree of flicker perceived was also known to vary with age; and, in general, people of 20 years of age were considered more likely to perceive high frequency flicker (50-100 Hz) while older people were more likely to be affected by low frequency flicker (below 20 Hz). The wide variability between people was known to be larger than these age effects, however, and it was possible to have old and young people with equal sensitivity. Brundrett's (1974) work had suggested that dissatisfaction with lighting was more often felt by those who could perceive flicker and that these people were also more likely to experience eyestrain or headaches.

Fluorescent lights on a 50 Hz supply were frequently reported to be the major source of flicker in modern buildings but there were four sources of this lamp fluctuation. Intensity and colour variations at 100 Hz were caused by the discharges within the lamp and by the differential time constants of the phosphors coating the tube. Neither of these were likely to be major causes of flicker however. Two sources of 50 Hz fluctuations were more likely to give rise to flicker. The first was due to the variations in discharge characteristics of anode and cathode such that an enhanced glow moved from end-to-end of the tube producing a 50 Hz variation in the last few inches at each end. Internal electrode shields could be used to reduce this phenomena. The second cause of 50 Hz light variations resulted from irregularities between the performance of the electrodes at each end of the lamp so that current flow in one direction was greater than that in the other direction. This effect was reported to increase with the age of the lamp, sometimes leading to significant amounts of flicker quite early in tube life (Brundrett, 1974). Proportionally small numbers of flickering tubes were considered unlikely to give rise to major problems within large installations however, (this observation appeared in the discussion to Brundrett's 1974 paper but no figures were quoted). It was also pointed out that replacement periods as short as 1 year were sometimes considered necessary (see B.6.5).

Flicker could be reduced by a number of techniques such as direct current supplies, multi-phase systems or high frequency supplies. A number of experiments were reported to have been carried out on different methods of reducing flicker (Brundrett et al., 1973; Segal, 1950; Rey and Rey, 1963). While some authors had described variations in the performance or fatigue of subjects, others had failed to find any effect. A major criticism of many of the experiments in this area was that too few subjects were used.

It was recommended that fluorescent tubes were chosen which gave rise to as few 50 Hz fluctuations as possible (e.g. by shielding the electrodes), that regular replacement was carried out, and that a 3 phase supply was used, each phase driving one third of the lights.

B.6.5 Maintenance and Replacement

The specified performance of a light fitting always referred to a new clean tube with a clean reflector, but performance was known to deteriorate markedly with age and dust deposits. Chapter 3 of the IES Code (1973) had been devoted to the question of how frequently light fittings should be cleaned and replaced. Figure 14, taken from the IES Code, shows how the light output of a fitting could decline by as much as 50% of its rated value after 3 years (9000 operating hours), even if cleaned every 1 year (3000 operating hours). These figures assumed that the lighted room was used for only about 10 hours per day.

If ANE 403 was to be operated for 24 hours a day in a windowless room this suggested that the light fittings should be cleaned at least every 4 months (3000 operating hours) and replaced at least every year (9000 operating hours). It was pointed out that the most suitable replacement cycle for a given light was usually recommended by the manufacturers and this could be less than, or greater than, that specified above.

The cost of replacing tubes individually would be very high, apart from any loss of performance resulting from the presence of badly flickering lights and from variations in light levels. Although it seemed wasteful to replace all the lamps regularly, the labour costs were usually much lower over a period of time and disruption of the operator's work would be reduced. However, it was also recommended that, if individual tubes filed or gave rise to unacceptable degrees of flicker, they should be replaced as necessary.

Planned maintenance of lighting was stressed to be necessary in order to ensure an even, high light level. By replacing tubes on a regular basis, uneveness of lighting, flicker and total cost was predicted to be reduced.

(By the same criteria it was pointed out that the television tubes of the datascreens also had to be cleaned, maintained and replaced at regular intervals to ensure that their performance would not fall below that recommended).

B.7. WORK LEVELS AND REST PERIODS

The task of operating ANE 403 was predicted to be visually demanding and to restrict physical movement to some degree, and therefore some consideration was given to enforced rest periods. Several studies had shown that fairly frequent rest pauses tended to improve total performance, even though less time may have been spent working. The purpose of the rest periods was to allow the operators to rest their eyes and to take some physical exercise; therefore it was thought to be advantageous if operators could be encouraged to move to a rest room during pauses rather than to sit at their desks. It was pointed out that, if movement to a rest room was to be encouraged, a minimum period of ten minutes would be necessary to make the journey worthwhile, giving the operator time to smoke a cigarette, drink coffee or do some physical exercises. In view of this fact, longer breaks, of ten minutes every hour for instance, were recommended in preference to more frequent but shorter breaks, of five minutes every half-hour, for instance. It was pointed out that experimental studies of work/rest periods might prove extremely useful, particularly if these could be combined with experimental measures of fatigue, eye strain, postural variations, etc.

Unless the operator found the physical exercise rewarding some way it was considered that the facilities were unlikely to be used. It was recommended that consideration be given to providing games requiring a degree of physical effort but which would at the same time be self-motivating. Games played against a machine with a published league table were thought to be ideal for individuals;

while badminton, table soccer or table tennis were recommended to provide entertainment for several players. It was considered that a call system might be needed, to enable supervisors to contact operators in the rest rooms if work-load was increasing or if the individual stayed too long. Cautionary note was also made of the fact that if drinks could be taken back to the console, provision had to be made on the desk for cups, waste buckets would be necessary, and the electronics had to be protected in case spillage occurred.

It was recommended that supervisors be made responsible for ensuring that compulsory rests were going to be taken and that late returning individuals would be recalled. In a tightly knit group social pressures were considered sufficient to maintain equity in relaxation periods.

Recommendation was made that the rest or exercise areas should not be a great deal more brightly lit than the working areas because this would introduce adaptation problems. This requirement ruled out physical games requiring high light levels (e.g. those with fast moving targets or hard heavy balls).

## B.8. THE ORGANIZATION OF WORKING GROUPS

The organization of telephone operators into working groups, and the specification of the roles of supervisors, was considered likely to prove at least as important as hardware/software characteristics and interface design. A proper analysis of the relevant factors was thought to be necessary before specific recommendations could be made in this area, but some of the relevant parameters were outlined, together with their possible effects on system performance.

Until this report appeared, it was apparent that the designers of ANE 403 had considered individual operators as identical, independent system units, who would differ only in the skills they would possess. As a result, the system had been designed so that operators could function independently, organized into functional groups only by the software of the computer system, albeit under the chief supervisor's guidance. The flexibility provided by the ANE 403 system could enable any operator to sit at any operating position and to be treated as a member of any skill group (e.g. French speaking operators). The flexibility was deliberately designed into the equipment to enable any telephone administration purchasing the ANE 403 system to organize the workforce in any of a wide variety of ways to suit its own working practices. However, the ergonomic consultants considered that certain patterns of work organization would prove satisfactory, others might have lead to major difficulties, and yet others might have lead to system collapse. In the next brief sections the questions of the organisation of language groups and of room organization are dealt with in greater detail.

B.8.1 Language Groups

As an example of some of the factors to be considered in work organization, three types of.possible organization of language groups were considered.

(a) Method 1

Operators could have been geographically distributed about a city, building or room, but treated as being members of functional groups for the purposes of call handling and supervision. Geographically defined groups could therefore have been comprised of members of various functional groups (e.g. speaking different languages to subscribers). In such an organization the system could have allocated any operator to another functional group without requiring them to move physically. Two types of supervision were possible under this scheme, both of which were considered likely to give rise to widespread problems. A supervisor could have been placed in charge of each geographic group but might not have possessed all of the necessary skills or might not have had access to enough information to deal with difficulties. Alternatively, each supervisor might have been made responsible for a functionally well defined group whose numbers were geographically separated. Supervisors operating this latter scheme were likely to find it very difficult to assess local situations accurately and therefore they might have been impaired in trying to solve problems effectively. In addition, remote supervisors would have had a lack of personal knowledge of operators, therefore utilizing them less effectively. The supervisor of distributed groups would have been limited to a scheduling role and would be unable to provide training, support and advice as effectively as someone placed locally. Finally, the supervisor would not have been able to identify the work group easily, and it would be impossible to hold them fully responsible for the standard of performance of the group. The problems outlined in this paragraph were predicted to be most strongly felt in widely distributed exchanges where, for example, operators might work in small offices located around the perimeter of a city close to plentiful housing.

(b) Method 2

In this situation the working groups would have been formed and reformed around sets of consoles to deal with changes in workload. Desks arranged in groups of 8 or 16 would become the temporary workplace of operators all working on the same types of call. As demand changed, operators would be asked to move to another work-station within a new group of people. This had the advantage that the functional groups could be seated together and could be locally supervised by someone with the necessary skills. Unfortunately, it was considered that individuals performing routine tasks would seldom like to use different workplaces and to sit with a variety of groups, and that they would usually prefer to feel that their immediate neighbours were kept fairly constant. A permanent place of work would also allow friendships to develop and would promote a feeling of group identity.

(c) Method 3

In this scheme groups would have been assembled in fixed geographic locations and re-allocated to functional groups as a whole rather than individually. In such an organization the members of each groups would all work together on identical types of calls with their own local supervisor, who would have the same range of skills which they possessed. This system was considered to reduce the use of the flexibility inherent in the ANE 403 system and it was noted that it may only be proved to be practical in large exchanges. However, the advantages of this type of system were manifold. Operators and supervisors would have first hand knowledge of local problems. A team identity could quickly develop, leading to increased co-operation and easier problem solving. Finally, the supervisor would know the skills, performance, habits and weaknesses of operators, facilitating their optimum use and enabling the supervisor to provide training, support and guidance. This organizational system was recommended.

B.8.2 Room Layout

In most traditional telephone exchanges operators' consoles were mounted in parallel rows, but on ANE 403 it was suggested that consideration should be given to other layouts. For instance, it was possible for groups of operators to be placed in consoles arranged to ensure that each member of the group faced inward. This was likely to provide a more satisfactory social environment, making the group more cohesive and able to cope with common problems. Local supervisors were more likely to find it easier to control geographically distinct groups, because visual contact with operators could be maintained.

The utilization of floor space by console arrangements was recognized to be an important economic consideration, but it was found possible to provide interesting and functionally useful layouts without necessarily occupying more total floor space.

### B.8.3 General Recommendation on Work Group Design

Socio-technical systems theory was well developed and at the time of the literature report the need for its application was by then quite evident in many large systems. A further investigation into the organisational issues raised above was strongly recommended.

## B.9 RECOMMENDED SYSTEM DESIGN

The following was intended to be a brief checklist of the recommendations made thoughout the literature review report. The footnotes indicate the extent of their adoption in the final design, sometimes altered as a result of the experiments reported in Chapters C and D, sometimes as a result of technical constraints and sometimes on ergonomics grounds after further consideration.

### B.9.1 Datascreen

| | | |
|---|---|---|
| Characters | Size: | 3.2 x 2.1 mm* (h x w) |
| | Spacing: | 0.6 mm* |
| | Line space: | 3.2 mm* |
| | Stroke width: | 0.4 mm non-interlaced* OR 0.3 mm interlaced+ |
| | Generation method: | 7x9 dot matrix* (with rounding+) |
| | Character design: | As in Figs. 5A and 5B.* |
| | Cursor: | Underline* with short vertical line at left hand end.+ Flashing at 3 Hz.* |
| | Brightness: | 120 and 150 $cd/m^2$.* |
| | Contrast ratio: | 10:1 in 300 lux.* |
| | Direction of Contrast: | Light on dark.* |

Footnotes

* Recommendations accepted and designed into the system.

\+ Recommendations not followed, usually after consultation with ergonomist.

| | | |
|---|---|---|
| Screen. | Size: | 10" screen with 125 x 175 mm* viewing area. |
| | Phosphor: | P31 or P39* |
| | Surround: | As in Fig. 12B. Matt and textured.* Colour - dark green.+ |
| | Viewing distance: | 500 - 600 mm* |
| | Screen treatment: | Contrast enhancing filter.+ Matt antireflection treatment.* |

B.9.2 Keyboard

| | | |
|---|---|---|
| Keys. | Design: | Rafi Tastaturen RS75* |
| | Colour: | Mid green with yellow characters or matt black with white characters.* |

Feedback. To special function keys (for exchange) and to editing keys, except cursor movement keys.+

Typomatic Functions: Typomatic functions on left and right cursor movements and space bar only*

Keyboard Case. Thin at front* and sloping at $20^{\circ}$ ‡ to the horizontal. Case to be free-standing.

B.9.3 Console

Design: As shown in Figs. 13A-C.*

---

Footnotes

* Recommendations accepted and designed into the system.

\+ Recommendations not followed, usually after consultation with ergonomist.

‡ Recommendations modified after further consideration in order to improve the ergonomics.

B.9.4 Dialogue

Power. Switch in headphone jackplug.+

Ticket.

a) Design as in Fig. 9.*

b) Cursor movements restricted to open fields.*

c) 2 brightness levels 120 $cd/m^2$ and 150 $cd/m^2$.*

d) Pictorial indication of exchange status.*

e) Field abbreviations to be changed as in Appendix 1.*

f) Reverse video for chief supervisor messages.*

Procedure.

a) Preformatted ticket to appear for vetting before accepted.*

b) Simple means of rerouting inappropriate calls.+

c) Inputs indicating the categories of calls to be individually accepted by the system in less than 2 s each.*

d) Operators should be able to book a call for a time interval not just a clock time.*

e) Automatic cursor movements should be considered.+

---

Footnotes

* Recommendations accepted and designed into the system.

\+ Recommendations not followed, usually after consultation with ergonomist.

B.9.5 Lighting

| | |
|---|---|
| Intensity: | 300 lux. § |
| Direction: | Vertically downward over screens.§ |
| Nature: | Artificial lighting - no windows.§ |
| Glare: | Glare index less than 16. § |
| Reflectances: | Matt or textured surfaces. § |
| Homogeneity: | Even light level but interest introduced by colour contrast. § |
| Supply: | 3-phases, independently distributed to lights.§ |
| Light fittings: | Fluorescent tubes chosen to reduce flicker problems.§ |
| Maintenance: | Cleaning every 4 months, replacement every year. (unless otherwise specified by manufacturer).§ |

B.9.6 Work Rest Periods

| | |
|---|---|
| Intervals: | 10 min. break per hour. § |
| Rest room: | Recommended. Relaxed atmosphere.§ |
| Physical exercise: | Self motivating games. § |
| Console: | Adequate protection from liquids if drinks taken to workplace.§ |

---

Footnote

§ Recommendations passed to buying administrations.

B.9.7 Organisation of Working Groups

| | |
|---|---|
| Consoles. | Regular console positions where possible.§ |
| Continuity. | Well established social and functional groups.§ |
| Supervisor. | Supervisor for functional groups to be geographically within groups where possible.§ |
| Load allocation | Supervisor to allocate work and leisure.§ |
| Layout of consoles | Consoles to be grouped rather than arranged in lines.§ |

---

Footnote

§ Recommendations passed to buying administrations.

# CHAPTER C - PHASE II

## EXPERIMENTS TO SELECT DATASCREENS AND TREATMENTS

### C.1 PREFACE

As a result of the literature review described in Chapter B it was possible to produce detailed recommendations for design, based upon ergonomics principles and published studies. However, as reported in Chapter B, there were some design conflicts which could not be resolved on the basis of a literature review alone, and a number of experimental studies were therefore proposed. One of the most important of these related to the choice and subsequent treatment of the datascreen. This study comprised Phase II of the research, and is reported in this Chapter.

The first stage of this phase was to narrow down the selection of datascreen designs and treatments which were later to be submitted to experimental test. At a series of meetings, representatives of LME Svensk Radio Ab (SRA - an Ericsson subsidiary responsible for the selection and design of the datascreen) and Loughborough University met several times to examine and discuss the various datascreens which might be employed in the electronic displays for ANE 403 and the various treatments which could be applied to the faces of the displays. The purpose of these meetings was to select, from available options, those screens and screen treatments which could be seen to be reasonable solutions for the operator terminal of ANE 403, and which were worthy of further investigation. In these meetings some options were rejected upon prima facie grounds and others were thought to be unsuitable on the basis of an informal experiment. In this informal experiment subjective assessments were made by those present.

As a result of this series of meetings a number of datascreens were rejected for a variety of reasons. Screens were rejected for poor image quality, 'spot wobble' or 'image swim'. Screen treatments were rejected because they led to blurring of characters or excessive reflections. Eight screen designs were finally accepted as meriting further investigation in fairly detailed experiments. These eight design solutions consisted of the combinations of two phosphors, P31 and P39, and four screen treatments. The four screen treatments were: a light matt surface directly applied to the screen, a matt neutral circularly polarising filter, a matt green circularly polarising filter, and a light-matt louvred filter manufactured by the 3M Company.

Phase II consisted of several experiments. The first two were directed towards selecting one screen design from the eight described above. The final two experiments were to submit the chosen design to a variety of lighting conditions, and to investigate any possible visual fatigue experienced by the subjects using the screen. In more detail, the four experiments were as follows:

Experiment 1  To reduce screen designs from eight to four in number, on the basis of the subjective assessments of a group of eleven people. The screens were viewed under an ambient light level of 300 lux.

Experiment 2  To evaluate the four screen designs selected in Experiment 1 and to select one of these screen designs on the basis of several measures:

(a) response time and errors on a task designed to measure relative legibility.

(b) subjective assessments of the screen characteristics.

(c) responses to a questionnaire to assess subjective visual fatigue.

This experiment, like Experiment 1, was carried out under 300 lux.

Experiment 3. The screen chosen in Experiment 2 was re-tested using the same methods used in Experiment 2 but under 500 lux of ambient illumination. The purpose of Experiment 3 was to ascertain if the screen could be used effectively under this higher lighting level.

Experiment 4. The selected screen design was presented to subjects under more ideal lighting conditions and at a workstation conforming to the specifications described in Chapter B.

In each of the four experiments described briefly above, the subjects were asked to perform a visually demanding task for four periods, each of twenty minutes duration. At the conclusion of each task period, the results were printed out, enabling performance to be monitored over a total period of eighty minutes of almost continuous screen use.

The common methodologies used in these experiments are described and discussed in Section 2 of this Chapter. Subsequently, each experiment is described in a separate section (C.3-C.6). In these sections the following aspects of each experiment are covered; the experimental method, the results, and a discussion. A discussion of the total study, and a description of the conclusions reach from Phase II, are presented in the final section of this Chapter (C.7).

## C.2 METHODOLOGIES

A wide variety of information, using different techniques, was collected in the four experiments which are the subject of the present Chapter. The information was collected by the techniques described in succeeding sub-sections.

### C.2.1 Subject differences (particularly those relating to the condition of subjects' eyes)

From the outset, it was anticipated that there were likely to be large intersubject differences in performance on the task, and that some of these differences might arise from the effects of extraneous variables, such as age, eyesight, intelligence, etc. To ensure comparability between subject groups assigned to each experiment and to the various conditions within each experiment, three alternative experimental techniques were considered:

- Each subject could have been exposed to each experimental condition.
- A large population of possible subjects could have been pretested to enable matching of subjects or to enable stratified samples to be selected.
- Subjects could have been randomly assigned to experimental conditions and subsequently tested to identify differences between the subject groups.

The first option entailed scientific and practical difficulties. The scientific difficulties arose from probable order effects (such as learning), the practical difficulties arose from ensuring that we could obtain subjects on several successive occasions. The second option could also not be used for practical reasons.

Insufficient resources were available from the company to identify a population of possible subjects before the experiment. In addition, there was insufficient time available to collect and analyse data on a large population of potential subjects from which representative samples could be selected for each experiment. Instead, each subject who participated in the present set of experiments was required to provide certain basic information on the day upon which they participated in the experiments. This information was used later to compare the composition of experimental groups, both within and between experiments.

Two techniques were used to collect information from subjects:

1. All subjects completed a 'pre-task questionnaire'. This is shown in Appendix 2. It will be seen that the questionnaire, which was given to all subjects before they participated in any of the experiments, consisted of four sections. Section A was concerned with the collection of general information such as the names of the subjects, their occupation, age, sex and so on. Section B was concerned with the condition of their eyes. Questions were asked relating to whether the subjects had any difficulties with their eyes, whether they had sight in both eyes, whether they were colour-blind and whether they used, or ought to have used, spectacles or contact lenses. Section C had to be answered only by those who wore spectacles, or who used contact lenses. Questions relating to the frequency of use of these optical aids and their precise form was posed, and questions were also asked on why it was that these optical aids were worn. Section D of the questionnaire asked questions relating to the general condition of the subjects' eyes. In

this section subjects were asked to state whether, when they read, they suffered from pain in the head, the neck, the shoulders or the back. They were also asked whether they suffered from any form of eye strain under various conditions.

2. An ophthalmic optician, Arne Elfström, carried out a series of simple visual screening tests using a Bausch and Lomb Orthorater. Using the apparatus, basic information on the condition of each subject's eyes was obtained. This information related to distance and close vision, and included measurements of both lateral and vertical phorias and of visual acuity. Measurements were also taken of depth and colour perception. Subjects who wore some form of optical aid, such as spectacles or contact lenses, were tested both with and without that aid. Appendix 3 shows the form completed by the ophthalmic optician after testing subject's eyes. The Appendix gives an English translation of the form and a copy of the Swedish original.

C.2.2 <u>Subjective impressions of the datascreen</u>

Subjective impressions have always been a useful means of comparing objects, actions and events. They have a particular value when a task is performed and no differences appear in terms of speed and accuracy between various experimental conditions. In these circumstances subjective impressions can be decisive in determining the relative merits of the conditions.

In this set of experiments it was considered that the following subjective impressions associated with various screen characteristics were important, based on the literature review and the opinions of the experts who participated in the first part of Phase II:

1. Amount of flicker perceived on the screen.
2. Clarity of the displayed characters.
3. Character "wobble".
4. The stability of the background surrounding the characters on the screen.
5. Degree to which light fittings were reflected by the screen surface.
6. Degree to which subjects could see reflections of themselves in the screen.
7. Attitude towards the colour of the screen characters.
8. General acceptability of the screen to the subject.

Most of the subjects were inexperienced in using data screens. Consequently it was felt that they would have difficulty in making absolute judgements of screen quality. It was therefore felt to be desirable to permit each subject to make comparative judgements of the screens wherever possible. The approach was to present screens simultaneously in pairs and for the subjects to state which one had "the most flicker", "clearest characters", and so on. Unfortunately, in certain experiments, with the limited numbers of screens and filters available, many pairs of design solutions could not be viewed simultaneously. After some experience with making paired comparisons subjects began to recognise the range of screen characteristics. As a result of this experience it was found that subjects could make fast, decisive and, as it proved, consistent absolute judgements of screen characteristics on five-point rating scales. The five-point rating scales used in the present experiments are shown in Appendix 4.

### C.2.3 Measures to assess visual fatigue

In order to assess the degree of visual fatigue brought on by using the screens for measured periods of time, it was necessary to devise tests which could be applied before and after exposure to the screens. Since visual fatigue had been found to be poorly defined in the literature, and since there were no universally accepted metrics to measure it, a subjective impression of visual fatigue as experienced by subjects was used. Subjective impressions were recorded by means of a questionnaire (the post-task questionnaire) shown in Appendix 5. This questionnaire was answered by subjects after they had completed their experimental sessions. The questionnaire consisted of three principal parts. The first simply recorded the names and other descriptive information of a general nature relating to the subjects. The second section was concerned with visual feelings or impressions as experienced by the subjects at the moment that they answered the questionnaire. The third section contained questions which related to the subjects' visual impressions as experienced during the experimental session.

Other objective measures of visual fatigue were considered and one was piloted. The degree of binocular fusion had been suggested as a measure of fatigue, since, after periods of time performing a task at relatively short viewing distances, the eyes were known to converge for a time when subsequently attempting to look at objects at infinity. The ophthalmic optician previously referred to in Section C.2.1 measured fusion at the same time as conducting the visual screening tests (also described in Section C.2.1). After periods of concentrated screen use, some subjects were re-tested for fusion. Because of difficulty in interpreting these data, results relating to the analysis of fusion have not been presented in this thesis.

### C.2.4 Visual task performance

In Experiments 2-4, subjects were required to carry out a task which was very visually demanding. The purpose of confronting subjects with this task was two-fold: to create a situation where differences (whether measured subjectively or objectively) between experimental conditions would be maximised, and to enhance the effects of passage of time upon visual fatigue.

#### Task

A computer game called HAMURABI was orginally suggested as a task for this set of experiments, but could not be consstructed in time and so a simpler task was devised for subjects to carry out. Subjects were seated at a suitably adjusted work station which incorporated a datascreen and keyboard. Subjects were presented with a series of single rows of digits.

Each row contained two sets of ten digits separated by 20 mm. The two sets of digits were symmetrically disposed about the mid-line of the datascreen . Figure 15 shows fourteen rows of digits and illustrates the task format. Each subject's task was as follows. The subject had to press one of two keyboard buttons. If the subject felt that the two sets of ten digits were identical then he was required to press a 'yes' button. On the other hand, if he felt that the two sets of digits within a row were not the same, then he was required to press the 'no' button.

The first row of digits appeared at the top of the screen. A short time after the subject had indicated his decision another pair of ten digit numbers appeared below the first. When the screen was full, having sixteen rows of digits on it, the next response caused the format to be erased and a new row of digits appeared at the top of the screen. After certain periods of time (varying with the experiment in question) the task was stopped, the subject was informed of the number of errors that he had made and the results were printed out. (See below and Figure 16).

Programmes for test material and data collection

A computer programme generated the rows of digits referred to in the previous section. The programme also recorded the subjects' decisions, stored their response times and produced 'hard' and 'soft' print-outs of performance for the experimenter at the conclusion of each experimental period.

A random number generator was used to create the pairs of ten digit numbers. In approximately 50% of cases the two numbers were identical and in the other 50% of cases one digit was different. A programme loop was used to measure the time from the presentation of a pair of numbers on the screen to the time when one of the two response buttons was pressed by the subject. This elapsed time constituted the response time.

Four categories of stimulus/response were possible. Two combinations represented correct decisions by the subject and two represented incorrect decisions. The possibilities were as follows:-

1. The pair of ten digit numbers was identical and the correct response button (yes) was pressed. On the print-out this was represented as A = B, OK.

2. The pair of ten digit numbers was identical and the incorrect response button (no) was pressed. This was represented on the print-out as A = B, Not OK.

3. The pair of ten digit numbers was not identical and the correct response button (no) was pressed. This was represented on the print-out as A ≠ B, OK.

4. The pair of ten digit numbers was not identical and the incorrect response button (yes) was pressed. This was represented on the print-out as A ≠ B, Not OK.

Examples of the frequency with which these different types responses occurred as given in Figure 16. This figure shows the results from subject 13 over two successive experimental periods.

As the experiment progressed, response times were accumulated for each stimulus/response category indicated above. Thus, it was a simple matter to calculate average response times for the two correct and the two incorrect decisions, knowing the number of stimulus presentations (i.e. rows of digits) that had been presented to each subject in the given time period. Additional data were also collected. For each stimulus/response category the square of each response time was taken and summed as the experiment progressed. This enabled the variances of response times to be calculated subsequently in a very simple manner.

The number of stimulus/response events, the total sum of the response time and the sum of the squared response times for each particular stimulus/response category for a particular subject could be displayed on a data-screen at the experimenter's console at the end of any experimental period. This information could be printed in hard-copy form onto a remote printer on demand from the experimenter's keyboard. A sample print-out of the results from two five-minute task periods for a single subject is shown in Figure 16.

Performance measures

The data collected for each subject during each task period (typically five minutes) were printed out as shown in Figure 16. This print-out enabled two performance measures to be calculated. These were: average response times for each type of response during each of the successive task periods and the error rate during successive task periods.

### C.2.5 Contrast and Brightness

In experiment 3, in order to assess the most effective contrast and brightness levels at different levels of illumination, it was necessary to collect two related pieces of information. These were:

1. The way subjects adjusted the brightness and contrast controls of the screen to their own preference.
2. The actual brightness and contrast levels corresponding to the settings made.

Two columns, ten digits wide, were displayed on each side of the screen midline. The left hand column was initially displayed at a relatively high brightness (but not one such as to make the characters blur). The right hand column was initially displayed at a relatively low brightness (but one at which the characters could be read). The brightness ratio between the left and right hand columns of digits was about 2:1. This brightness ratio had been chosen by engineers to distinguish between newly entered characters on the one hand, and field labels and previously entered characters on the other (in the proposed ANE 403 "call ticket"). The experimenter set the contrast control, which determined the brightness ratios between the characters on the screen and the background, at four different levels. At each of these levels, subjects adjusted the brightness control so that the character brightnesses in both columns were at a preferred level. After making brightness control adjustments at predetermined contrast control settings subjects were permitted to adjust both controls simultaneously to obtain preferred screen and character brightnesses. These procedures were carried out under two levels of ambient illumination; 300 and 500 lux.

Character and background brightnesses associated with each setting of the contrast and brightness controls were measured in $cd/m^2$ using a Hagner photometer. These measures of luminance were made on a special screen format where the characters were removed and replaced by broad bands of light at the same luminance level as the characters they replaced. Thus readings (for each control setting) were taken on three bands of light on the screen: one corresponded to the high brightness characters, one to the low brightness levels, and one to the background.

## C.3 EXPERIMENT 1 : REDUCTION OF SCREEN OPTIONS BY SUBJECTIVE ASSESSMENT

Before the present experiments began, a number of experts examined a number of datascreens and screen treatments to determine, on prima facie technical, ergonomic and cost grounds, what design solutions appeared to merit detailed experimental consideration. As a result of these expert investigations eight datascreen designs were put forward for testing in Experiment 1. These eight solutions were produced by the combination of tubes coated with P31 and P39 phosphors and treated in one of four ways. These eight options have been already described in the third paragraph of section C.1 and are further described in the part of section C.3.1 which deals with apparatus (see below).

### C.3.1 Experimental Method

#### Experimental Design

Eleven subjects each viewed all eight datascreen designs, first making comparative judgements by paired comparisons and, later completing a series of five-point absolute rating scales.

#### Subjects

Eleven female telephone operators, selected at random from those employed at L.M.E.,participated in this experiment.

#### Apparatus

Four video monitors were used in this study, two each with the P31 phosphor and two each with the P39 phosphor. For each type of phosphor one screen had a matt finish applied directly to the screen and the other had an untreated screen with a glossy surface. The following three filters were used in conjunction with both of the glossy screens:

1. Green circularly polarising sheet 1.83 mm thick with a matt finish on its outer surface. (Supplied by the Polaroid Corporation - reference No. HGCP 21).

2. Neutral circularly polarising sheet 0.76 mm thick with a matt finish on its outer surface. (Supplied by the Polaroid Corporation - reference No. HNCP 37).

3. A louvred green "light control film" with a "very light matt" finish on the outer surface. (Product supplied by the 3M Corporation who recommended this finish for datascreen use.)

The circularly polarising sheet was claimed to improve contrast enhancement by causing ambient light passing through the filter from the outside to be totally absorbed between the filter and the screen surface (see Appendix 6). The "light control film" incorporated tiny louvres and was said to achieve contrast enhancement by selectively absorbing light incident on the filter at angles well outside $90^{o}$ to the filter surface (see Appendix 7).

As already stated there were eight experimental conditions as follows:-

| Phosphor | TREATMENT TO TUBE FACE | | | |
|---|---|---|---|---|
| | | Filters added to a glossy screen | | |
| | Matt finish applied directly to tube face | Green polaroid matt finish | Natural polaroid matt finish | 3M light control filter (light matt finish) |
| P31 | A0 | A1 | A2 | A3 |
| P39 | B0 * | B1 | B2 | B3 |

* The matt surface applied to this screen was not a standard finish and was poorly applied.

The monitors were each fed by the same video signal, which produced a "call ticket" on the screen (see Figure 9). The tickets had information at two levels of brightness, the field labels in low brightness and a single line of text in high brightness.

The four monitors were placed on a two-level desk, such that the two monitors at the rear were higher than the two at the front (see Figure 17). The position of the monitors was changed for each subject and the screens were labelled for identification purposes by the experimenter. The three filters were mounted on card holders which simultaneously cut down the screen size to the ANE 403 "call ticket" size and provided a carrier for the filters to prevent them from being marked by fingerprints. The card holder also enabled the filters to be affixed to the monitor screens. The filter holders were also appropriately labelled.

300 lux of ambient lighting was provided on the top surface of the monitors by a total of six fluorescent tubes (one tube was switched on in each of six pairs). The experimental room was small compared with the types of offices which might be built for the operators of ANE 403. Consequently no luminaires could be fitted on the ceiling sufficiently behind the operators to cause potential reflections on the screen. Accordingly, to simulate a larger room with poorly designed light fittings,a large mirror was mounted on an easel behind the subject, such that the light from the luminaires was reflected to shine over the subjects' shoulders and onto the screen (see Figures 18 and 19) Thus,the lighting conditions were made much more severe than they would be in a well designed operators' room,in order to subject the screens to a very severe test.

Subjects were seated on an adjustable swivel chair in front of the work station on which the monitors stood. The height of the work station was not adjusted for each subject because the four monitors were very heavy and because adjusting height would have altered the light levels at the monitors. However, to ensure that all subjects had their heads at the same height and position the swivel chair was appropriately adjusted.

Procedure

Two sets of judgements were made by each subject. The first involved making comparative judgements between pairs of screens presented simultaneously. The second set involved absolute judgements, (see section C.2.2 for a fuller description of these judgements). Comparative or absolute judgements were made on the eight screen characteristics described in section C.2.2.

To be more precise, for each screen characteristic each subject made paired comparison judgement on the following screen pairs:

A0/A1, A0/A2, A0/A3

B0/B1, B0/B2, B0/B3

A0/B0

Absolute judgements on a five point scale were made for each of the screen characteristics, by each subject, on each of the following screens.

A1, A2, A3, B1, B2, B3.

This approach was forced on the author because there were no duplicates of the filters to be added to the glossy screens. Nevertheless, the present approach permitted comparative judgements to be made for a given phosphor, between a matt finish applied directly to the tube face and various filters added to a glossy face. Comparative judgements were also possible between phosphors for screens with a matt finish applied directly to the tube face. As previously indicated, absolute judgements were made on all the screens which were glossy and which were fitted with additional filters. These absolute judgements were made after subjects had carried out comparative judgements and were, perhaps, more sensitive to desirable and undesirable screen attributes.

### C.3.2 Results

Paired comparisons

The detailed results of the paired comparison tests are given in Table 1, which indicates the preferences of the eleven subjects on each pair of screens presented to them for each of the eight screen characteristics. Table 2 summarises the results, showing only the preferences which were statistically significant.

Examining first the preferences between the screens with P31 phosphors (A0 v. A1, A2 and A3), it is clear that the matt green circularly polarising filter over the glossy screen (A1) was significantly preferred to the matt surfaced screen (A0) on flicker, character sharpness, colour and overall acceptability.

The other filters in combination with the P31 phosphor (A2 and A3) were judged to be not significantly different from the matt surfaced screen (A0), except that combination A3 led to significantly more reflections from the operator than did A0.

Studying the results for the screens with P39 phosphors (BO v. B1, B2 and B3) it is apparent that the green polaroid screen of B1 was only better than the matt screen of BO on character wobble. The matt screen, BO, was better than screens B2 and B3 in terms of reflections of lights and better than B3 in terms of reflections of the operator.

When comparisons were made between the matt finished screens for the P31 and P39 phosphors (AO v BO), the P39 phosphor was significantly preferred on flicker and character wobble. The P31 phosphor was significantly preferred in terms of character sharpness.

Subjective absolute ratings

Table 3 shows the average absolute ratings given by subjects to the six combinations of three filters and two phosphors (i.e. conditions A1, A2, A3 and B1, B2 and B3). These combinations were judged to be significantly different on all the characteristics except colour and overall acceptability.

As might be expected, the combinations with a P39 (long persistence) phosphor screen (i.e. B1, B2 and B3) were seen to flicker less and to have more stable characters and a steadier background than the combinations with P31 (short persistence) phosphor screens (i.e. A1, A2 and A3). It must be noted, however, that the matt green polaroid filter markedly attenuated these disadvantages of the P31 phosphor (see results of A1 v A2 and A3).

In terms of character sharpness it is clear that the green polaroid (A1 and B1) and the 3M filters (A3 and B3) provided much sharper characters than did the neutral polaroid filter (A2 and B2).

The results of the judgements on reflections (of light fittings and of the operator) are also very clear-cut, although none of the combinations of screens and filters gained high scores because the test had purposely been made severe, as has already been described. As might be expected the most matt filter, the green polaroid (A1 and B1) gained the best average ratings for the screen characteristics relating to reflections.

As stated previously, on colour and general acceptability the differences between the ratings given to the screens and filters were not statistically significant.

Table 4 gives a simplified picture of all of the above results. The data was simplified by stratifying the average ratings (as shown at the foot of the table) thus enabling the advantages and disadvantages of the screens and filters to be assessed. In Table 4 it is clear that, in addition to information already presented on absolute ratings, the P39 phosphor with a matt green polaroid filter (B1)was very much liked. It is closely followed by the P31 plus matt green polaroid filter (A1) and the P39 plus 3M filter (B3).

### C.3.3 Discussion

It is clear from this brief experiment that specific screen and filter combinations gave rise to statistically significant differences in judgements of a number of screen characteristics. These differences enabled a smaller group of screens and filter combinations to be selected for further and more detailed experiment. Moreover, it is possible to make more general statements based on the results of this experiment as follows:

1. The green polaroid filter attenuated flicker markedly particularly on the P31 phosphor (see Tables 1 and 3).

2. The green polaroid filter and the 3M filter gave sharper screen characters than did the neutral polaroid filter (see Table 3).

3. The matt finish of the green polaroid led to less reflections of light fittings and of the operator than did the other filters (see Table 3).

4. Although the green polaroid filter improved the characteristics of the P31 phosphor the other filters were generally less preferred than the matt finish applied directly to this screen (see Table 1).

5. The addition of filters to the P39 phosphor always led to improvements over the matt screen except that in some cases reflections were increased (see Table 1).

Although these improvements were not statistically significant (by the binomial test) they were large and consistently in the same direction (see Table 1). In this context it is worth noting that the matt finished P39 monitor used had a number of technical faults and had a poor matt finish which was not a standard treatment.

6. P39 phosphor was perceived to flicker less, gave less character wobble and had a more stable background than did P31 phosphor (see Table 3).

Based on this evidence, which effectively summarises that presented earlier, four screens were selected for Experiment 2:

The matt green polaroid filter applied to the P31 and P39 screens had consistently proved to get good ratings on most of the screen characteristics. Accordingly conditions A1 and B1 were chosen for further experimentation.

The P31 matt screen (A0) was chosen because it had proved reasonably good and because it was wished to assess the possible effects of flicker on operator eye fatigue. Accordingly condition A0 was chosen for further experiment.

The P39 plus 3M filter (B3) was chosen because it had consistently gained higher (but not statistically significant) ratings than the P39 matt screen, except on reflections, which it was thought could be reduced in a proper installation. Condition B3 was therefore carried forward

## C.4 EXPERIMENT 2: SELECTION OF ONE DATASCREEN DESIGN FOR ANE 403 BASED UPON TESTS OF LEGIBILITY, SUBJECTIVE ASSESSMENT AND VISUAL FATIGUE

As a result of Experiment 1, only four datascreen designs were considered worthy of further study. In Experiment 1 and in the prior expert evaluations, screens had been selected entirely on the basis of subjective assessments. The purpose of Experiment 2 was to select a single datascreen design from four (AO, A1, B1, B3) on the basis of three sorts of evidence:

1. Relative legibility as measured by a task requiring digits to be read from the screen.
2. Subjective judgements of visual fatigue after using the datascreen for some time.
3. Subjective evaluation of the datascreens on the basis of eight screen characteristics.

### C.4.1 Experimental method

#### Experimental design

Subjects were divided into four groups. Each group was assigned to only one of the four datascreens. Thus each subject only used one screen and each screen was used by a separate and independent group of subjects. Such a design eliminates possible order and transfer effects which can contaminate results when each subject is assigned to a number of experimental conditions which he experiences in rapid succession. On the other hand, the method of allocation of subjects such that each subject experiences only one experimental condition leads to the confounding of individual differences and experimental effects. However, this is a rugged approach, because if individuals are broadly similar and significant effects occur then the effects are likely to be important.

## Subjects

Thirty-nine female subjects took part in Experiment 2, none of whom had participated in Experiment 1. It had been hoped that telephone operators could be obtained, but, due to lack of their availability and to time pressures, other subjects were taken who came from a variety of clerical jobs. The subjects were assigned to four groups, with 10, 9, 10 and 10 subjects in each. Originally there were ten subjects in each experimental group, but unfortunately one subject dropped out at short notice and could not be replaced. The original plan called for the subjects to be randomly allocated to the four groups in such a way that the groups were rated on certain important characteristics (such as visual acuity). Unfortunately, a proper balancing of subject groups eventually proved impossible due to the difficulty of obtaining data on the subject groups well in advance of the experiment so that matching could be performed. Subjects were therefore allocated to the four groups solely by their order of arrival. However, a number of important measurements were taken to enable subsequent comparisons of the constitution of the subject groups to be made, if required. The groups were numbered 1-4 to correspond with the numbering of the screens they used (see below).

## Apparatus

The four subject groups used the following in Experiment 2:

Group 1. A P31 phosphor tube with a matt finish applied directly to the screen (A0 of Experiment 1).

Group 2. A P39 phosphor tube with a glossy screen, over which was fitted a light matt 3M light control film filter (B3 of Experiment 1).

Group 3. A P39 phosphor tube with a glossy screen, over which was fitted a green circularly polarising filter with a matt finish (B1 of Experiment 1).

Group 4. A P31 phosphor tube with a glossy screen, over which was fitted a green circularly polarising filter with a matt finish (A1 of Experiment 1).

The screens were individually mounted on a workstation with a fixed level support surface for the monitor. There was, however, a surface adjustable in height for the keyboard. The monitors could all be tilted back by up to 10°, either by using an integral tilt mechanism, fitted to two of the monitors, or by using a specially constructed base plate, (see Figure 20 for a general view of the workstation).

The visual performance task which the subjects carried out using the screen was described in Section C.2.4. The apparatus needed to generate the task to collect data on performance was also described in Section C.2.4. By way of illustration, Figure 21 shows a close-up of a datascreen with two columns of 10 digit numbers. It will be recalled that a row of digits appeared at a time. Each row was divided into two sets of ten digits and the subject had to state whether the right-hand set was identical to, or different from, the left-hand set. Figure 22 shows a subject seated, at her task with the experimenter facing her. Figure 23 shows a close-up of the experimenter's screen. On it are portrayed data relating to the subject's performance over the previous five minute period, (see also C.2.4, and Figure 16, where a hard copy of similar data is portrayed).

The workstation, at which the subjects sat, was illuminated by six fluorescent tubes, providing a total of 300 lux in the horizontal plane at the screen surface. A large mirror, mounted upon an easel behind the subject, was angled such that light from one of the luminaires was reflected over the subjects' shoulders to strike the datascreen normally (i.e. at 90 degrees to the screen surface). The position of the mirror was identical to that used in Experiment 1 (see Figure 18) and is shown in graphical form in Figure 24.

Other apparatus used included the following:-

(a) a pre-task questionnaire (see section C.2.1 and Appendix 2).

(b) a Bausch and Lomb Orthorater (see section C.2.1 and Appendix 3).

(c) a series of rating scales to evaluate screen characteristics (see section C.2.2 and Appendix 4)

(d) a post-task questionnaire (see section C.2.3 and Appendix 5).

The purposes for which these various instruments were used are described in the Sections referred to above.

Procedure

The experiment took place in three stages. In the first stage the company optician carried out a visual screening test on each subject (see Section C.2.1 and Appendix 3) and subjects completed the pre-task questionnaire (see Section C.2.1 and Appendix 2). In the second stage, subjects performed a visually demanding task for six five minute periods with a break of thirty seconds between each period. The task is described in detail in Section C.2.4. The third and final stage consisted of subjects completing two questionnaires. In one, they recorded their subjective impressions of eight screen characteristics (see Section C.2 and Appendix 4). In the other they reported their subjective impressions of visual fatigue (see Section C.2.3 and Appendix 5).

The total time each subject was involved was approximately one hour.

## C.4.2 Results

Pre-task tests: Subject differences

As mentioned in Section C.4.1, the subjects were given a visual screening test and a pre-task questionnaire before they carried out the main experimental task. Unfortunately, as was also mentioned previously, it proved impracticable to use the results from the questionnaire and the screening test to balance the subject groups assigned to the four datascreen designs. However, the results were used as a *post hoc* means of comparing the construction of the subject groups.

The results from the visual screening test are given in Table 5. At this stage the subject groups had not used any screens and they are therefore referred to by Group number. The screens they eventually were to use are indicated for information. In the first section (tests 1 - 4) the numbers of subjects wearing different types of spectacles are listed. It should be noted that some subjects had more than one type of spectacles. It is readily apparent that the numbers of subjects within the groups who wore 'far' or 'near' correction spectacles differed little. More subjects, however, in groups 1 and 2 wore bifocal spectacles than did subjects in groups 3 and 4.

In the second and third sections (tests 5 - 9 and 10 - 14) the incidence of far and near visual problems are noted. The number of subjects in each group who fell just below the minimum standards required for Clerical and Administrative work, on the Purdue University Standard 1, (Bausch and Lomb Incorporated Vision Tests), are listed with the suffix 'Y' (Yellow band). The numbers of subjects who had "seriously lowered vision" (as defined by the Purdue Standard) are listed with the suffix 'R' (Red band). It is apparent that, while the incidence of mild and severe problems in far vision showed no clear differences between groups, there were higher incidences of mild problems in near vision in groups 1 and 2 than there were in groups 3 and 4.

The results from the pre-task questionnaire are listed in Table 6. The answers to question A3 indicate that the subjects in groups 1 and 2 were, on average, older than those in groups 3 and 4. (This difference was significant at the 0.01 level, analysed by a Kruskal-Wallis one-way analysis of variance.) As in the visual screening test (see above) the higher incidence of spectacle wearing in groups 1 and 2 is again noted (B4). No easily interpretable pattern emerges from the results to questions C1 to C7 inclusive. No major differences can be seen in the frequency with which subjects in the different groups reported the visual fatigue symptoms listed in question D1. Slightly more people in groups 3 and 4 reported having headaches often or seldom (D2a), and more of these subjects were also willing to attribute the headaches to eyestrain (D2b). Subjects in groups 3 and 4 also reported more eyestrain when driving (question D4d) and a greater degree of irritation to lights, cigarette smoke and dry air (questions D5 to D7 inclusive).

The results of the visual screening test and the pre-task questionnaire may be summarised as follows. Subjects in groups 1 and 2 were older, more wore bifocal spectacles, and more had visual problems detected in the visual screening tests than subjects in groups 3 and 4. Groups 3 and 4, however, contained more subjects who reported visual problems, more who reported headaches and more who were willing to associate these headaches with visual fatigue than groups 1 and 2. Furthermore, more subjects in groups 3 and 4 reported eye irritation from lights and cigarette smoke than did the subjects in groups 1 and 2.

Visual task performance

During the main experimental task (described in detail in Section C.2.4) the computer programme which generated and displayed the test material also stored data on the subjects' responses.

It will be recalled (Section C.2.4) that the following:- combinations of stimulus and responses were possible:-

Correct response 1 - The two 10-digit numbers were correctly reported to be identical (A = B, OK).

Correct response 2 - The two 10-digit numbers were correctly reported to be different (A ≠ B, OK).

Incorrect response 1 - The two 10-digit numbers were falsely reported to be identical (A ≠ B, Not OK).

Incorrect response 2 - The two 10-digit numbers were falsely reported to be different (A = B, Not OK).

For each of these combinations of stimulus and response the following data were collected for each subject, for each experimental condition and for each five-minute period.

1. Total number of times that this stimulus response combination occurred.
2. The sum of the reaction times (in milliseconds).
3. The sum of the squares of each response time.

See Figure 16 for an example of the printout in hard copy form. It can be seen from the figure that in Period 1 (of five-minutes duration) subject 13 made 15 correct responses to identical numbers (i.e. A = B, OK) in a total response time of 105089 milliseconds (i.e. 7.01s average response time). The sum of the squared response times was 113667389 milliseconds, indicating that the standard deviation of the response times within that five-minute period was 5.1s, calculated from the formula:

$$sd = \sqrt{\frac{\Sigma x^2}{N} - \frac{(\Sigma x)^2}{N^2}}$$

Percentage error rates could be calculated from the stored data according to the following formula:-

$$\text{Percentage error} = \frac{\text{(Number of incorrect responses)}}{\text{(Number of presentations)}} \times 100$$

Percentage error rates for each subject were calculated on the basis of all six five-minute periods taken together (i.e. thirty minutes in all). This was done:

(a) because it seemed implausible that error performance would vary substantially over so few periods of time,

(b) because it was considered that there were insufficient stimulus presentations within each five-minute period to obtain a representative sample of responses for an accurate assessment of error rate in that time period.

Although it would have been possible to separate the error rates for the two types of error (A = B, Not OK; and A ≠ B, Not OK), the extremely low rate of "A = B, Not OK" errors for some subjects did not permit accurate estimates to be made of the rate for this type of error.

Table 7a shows the average percentage error rates for the subjects in each of the four groups. Though the different datascreen/treatment combinations appear to produce substantially different mean error rates (the percentage error rate for screen B3 is more than twice that for screen B1), analysis of variance shows that there is no significant statistical difference between the screens in terms of this measure.

The average time it took each subject to make both correct and incorrect responses was calculated from all the responses made within all six five-minute periods. Average times for each subject group are shown in Table 8a. These response times appear to differ little. This is confirmed by the analysis of variance on the mean times. There is no significant statistical difference between the different datascreen/treatments in terms of this metric. It is interesting to note, comparing Tables 7a and 8a, that the subject group which obtained a relatively low error rate also took relatively longer to respond (i.e. Group 3).

To discover if there were any significant differences between the datascreen/treatment combinations in terms of subjects' times to make correct responses or false responses, statistical analyses were carried out. These are shown in Tables 9 and 10. No significant differences between the datascreen/treatment combinations were apparent for either of these metrics. It will be noted that the degrees of freedom in the Analysis of Variance summary in Table 10 are lower than in Tables 7 - 9. This discrepancy arises because some subjects made no errors during any of the six five-minute periods and consequently provided no data for this analysis.

Post-task tests: Visual fatigue and screen characteristics

After performing the experimental task, subjects were asked to complete a post-task questionnaire (see Section C.2.3 and Appendix 5) and to rate the datascreen on eight important characteristics (see Section C.2.3 and Appendix 4)

The results from the post-task questionnaire are shown in Table 11. In general there were few, if any, differences between the results for the groups who had used the different screens; although there was a slight tendency indicated for subjects who had used screens A0 and A1 to report symptoms of fatigue more frequently (see totals for question B1), than subjects who had used screens B3 and B1. It may be significant that screens A0 and A1 were each fitted with the harder P31 phosphor which tended to be seen to flicker more and to give rise to more character wobble than screens B3 and B1 which were coated with P39 phosphor (see below).

The average ratings given to the eight screen characteristics are shown in Table 12. As in Experiment 1, subjects again awarded significantly different ratings to screens on certain characteristics. On almost all the scales, screen B1 was preferred, being significantly better on flicker and character wobble. It must be pointed out, however, that screen B1 was significantly worse than screen A0 on the scale referring to the reflection of lights. This result was unexpected because screens B1 and A1 were fitted with matt green circularly polarising filters designed to enhance contrast and to reduce specular reflections from light fittings. Screen A0, however, was treated by a matt finish applied directly to the screen surface. It appeared that the heavily matt surface of the filter fitted to screens B1 and A1, which it will be remembered was only bent in one plane, diffused the light over a wide area of the screen, thereby obscuring the whole display. The curved surface of the exposed matt screen of screen A0 appeared to minimise the spread of the reflections, focussing them on a small area of the display, such that head movements were sufficient to move their positions on the screen and thereby overcome their masking effect.

It will be remembered that the lighting conditions were particularly severe in this experiment (with a large mirror to reflect a light fitting, thereby simulated a badly designed luminaire). Under better lighting conditions, and with a more effective workplace design, the problems should be minimised (see the results of Experiment 4). More effective bonding of the filter to the screen surface might also help to reduce the effects of reflections.

### C.4.3 Discussion

There were no significant differences between the performance of the subject groups who used the four different datascreens. This indicates that the differences between subjects were much larger than any relative differences in legibility between the screens. Such a finding is not, perhaps, very surprisingly, when it is remembered that the screens used in this experiment had already been selected by experts and the subjects in Experiment 1 as potentially useful designs for ANE 403.

Despite these insignificant differences in task performance, subjects consistently judged the screens differently on a number of other important characteristics: flicker, reflections of light, etc. Screen B1, consisting of a monitor with a P39 phosphor and a glossy screen, over which was fitted a matt green circularly polarising filter, was chosen as the best design on the basis of these subjective judgements. One important proviso over the use of screen B1, however, was that the lighting conditions in installations would have to be carefully designed to reduce the possibility of reflections of lights in the screen surface.

After completing the main experimental task subjects completed a post-task questionnaire which enabled them to report the degree of fatigue they experienced. Few differences were found between the replies obtained from the groups of subjects who used the different screens.

## C:5 EXPERIMENT 3 : COMPARISON OF THE PERFORMANCE OF A DATASCREEN FOR ANE 403 UNDER 300 and 500 LUX AMBIENT ILLUMINATION

As a result of Experiment 2, a monitor with a P39 phosphor covered with a matt green circularly polarising filter (B1) was chosen as the most effective datascreen design for ANE 403. Experiment 2 was conducted under 300 lux of ambient illumination, since this was the maximum level recommended in Phase I, (see Chapter B). In the experimental programme drawn up for Phase II it had been intended that the effects of ambient illumination would be studied. The original plan involved investigating the level of 150 lux (the lowest level recommend by Phase I), against 300 lux, the maximum recommended. It was subsequently decided that 150 lux appeared to be too gloomy for a large office and it was considered that the screen might be adequately legible under 500 lux, illumination conditions which might provide a more cheerful and pleasant environment.

The purpose of Experiment 3 was to repeat the tests used in Experiment 2 using the chosen datascreen (B1) and ten more subjects, but under an illumination level of 500 lux. Thus, the results for the chosen datascreen already obtained in Experiment 2 under 300 lux could be directly compared with the results obtained in Experiment 3 under 500 lux.

The same workstation, light fittings and stringent lighting conditions (i.e. maximising reflections of the light fittings) were used in Experiment 3 as had been used in Experiment 2.

In Experiment 3 the brightness and contrast levels of the datascreen preferred by subjects under both 300 and 500 lux lighting conditions were measured. The positions of the control knobs giving the preferred conditions were recorded so that engineers could specify the likely life of the datascreen at these settings.

## C.5.1 Experimental Method

The methodology employed in Experiment 3 was similar to that used in Experiment 2. The only difference was that some additional information was collected in Experiment 3.

### Experimental Design

The datascreen chosen from Experiment 2 (B1) was retested by a totally new group of experimental subjects (Group 5) under conditions identical to those used in Experiment 2 except for a higher ambient lighting level (500 lux rather than 300 lux). The same experimental methods were used in Experiments 2 and 3, but some additional information on brightness and contrast control settings was collected in Experiment 3. Thus the results obtained in Experiment 3 could be directly compared with the results collected in Experiment 2 for the chosen data screen (B1).

### Subjects

The ten female subjects who took part in Experiment 3 had not been used in any of the previous experiments. They came from a variety of clerical jobs in LME. As in Experiment 2, data were collected on subjects' eyesight and proneness to symptoms of visual fatigue, but, unfortunately, these data could not be used to match the subject groups, and were only used for post hoc comparisons.

Apparatus

The datascreen used in this experiment was that used by Group 3 in Experiment 2; it comprised a monitor with P39 phosphor and a glossy screen surface, over which was fitted a green matt finish circularly polarising 'Polaroid' filter 1.83 mm thick (code B1). The filter was bent to touch the screen all the way across but would not bend to touch the screen at the top and bottom.

As in Experiment 2, the datascreen was mounted on a workstation with a fixed screen support but with an adjustable keyboard height (see Figure 20). The performance task which the subjects carried out using the datascreen was identical to that used in Experiment 2 (see section C.4.1) and described in detail in section C.2.4. The task material and data collection techniques were also identical to those used in Experiment 2 (see section C.2.4 for details).

The light level of 500 lux used in Experiment 3 was obtained from six double-tube fluorescent light fittings, each tube consuming 40W of power. The arrangement of the lights is shown in Figure 24. To produce 500 lux, two tubes in each of the four fittings, A, B, C, D were switched on and one tube in each of the fittings E and F were switched on (see tube numbers indicated in Figure 24). The same large mirror described in Experiment 2 (section C.4.1) was used to provide a severe test for screen reflections, (see Fig.18 and Elevation in Fig.24).

As in Experiment 2 the following apparatus was also included:-

(a) a pre-task questionnaire (see section C.2.1 and Appendix 2)

(b) a Bausch and Lomb Orthorater (see section C.2.1 and Appendix 3

(c) a series of rating scales to evaluate screen characteristics (see section C.2.2 and Appendix 4).

(d) a post-task questionnaire (see section C.2.3 and Appendix 5).

The purpose for which these various instruments were used are described in the sections referred to above.

In addition to the apparatus listed above, which is identical to that used in Experiment 2 except for the higher lighting level, the brightness and contrast control knobs were fitted with numerical dials to indicate the position of their settings. These dials were arbitrarily numbered 0-11 on linear interval scales.

Procedure

The experimental procedure was identical to that used in Experiment 2, except that only one group of subjects was tested and some additional measures were made in the final post-task phase. As in Experiment 2 there were three experimental phases.

In the first phase, the company optician conducted a visual screening on each subject, who then completed a pre-task questionnaire. During the second phase subjects performed a visually demanding task for six five-minute periods with a break of 30 s between each period. As in Experiment 2 subjects were told during each break of the total number of errors they had made in the previous five minute period. The third phase consisted of subjects completing two questionnaires. In one they recorded their subjective impressions of the screen on eight screen characteristics (see section C.2.2. and Appendix 4). In the other questionnaire they reported their subjective impressions of visual fatigue etc. (see section C.2.3 and Appendix 5). During this final post-task phase, subjects were asked to carry out tasks in Experiment 3 which were additional to those performed during Experiment 2. Described in detail in section C.2.5, these tasks involved the subjects adjusting the brightness and contrast control settings of the screen to obtain their preferred levels for those values. To provide complementary information for Experiments 2 and 3 these preferences were obtained under both 300 lux and 500 lux ambient lighting levels.

### C.5.2 Results

#### Pre-task tests: Subject differences

As discussed in section C.4.1 the subjects were given a visual screening test and a pre-task questionnaire before they carried out the main experimental task. Unfortunately as in Experiment 2, it proved impracticable to use the results from the questionnaire and the screening test to match subjects. The results of these pre-task questionnaires were, however, used for a post hoc comparison of subject groups 3 and 5. Subject group 3 had used the chosen datascreen under 300 lux in Experiment 2, and subject group 5 used the same screen under 500 lux in Experiment 3.

The results from the visual screening test are given in Table 13. It can readily be seen from this table that more subjects in Group 3 (from Experiment 2: condition B1) had poor acuity (for both far and near vision) than in Group 5 (from Experiment 3: condition B1). The results from the pre-task questionnaire are given in Table 14. Group 3 had a higher average age and contained more spectacle wearers, than Group 5. Answers to questions relating to eyestrain in specific situations elicited answers which seem to indicate that more subjects in Group 3 experienced a greater degree of eyestrain when reading, watching television and watching cinema. The smaller proportion of drivers in Group 3 than in Group 5 who reported experiencing eyestrain when driving may be explained by other factors than proneness to eyestrain (driving frequency, type of driving etc.). More Group 3 subjects reported irritability of their eyes to lights and to cigarette smoke.

Visual task performance

As described in section C.4.2 the error rate and the response times (for correct and incorrect responses) were collected for each of six five-minute periods for each subject.

As explained in section C.4.2,the average percentage error rates were calculated for each subject over six periods and then averaged over subjects to provide the means shown in Table 15a. The mean percentage error rate under 500 lux was almost twice that under 300 lux, but this difference (4.39 and 2.29) was not statistically significant, however, as the analysis of variance in Table 15b shows. Tables 16-18 inclusive show the statistical analysis of the results for different types of response times. For no metric were differences significant between the 300 and 500 lux conditions.

<u>Post-task tests: Visual fatigue, screen characteristics and brightness/contrast settings</u>

After performing the experimental task, subjects were asked to rate the datascreen on eight important characteristics (see section C.2,2 and Appendix 4), and to complete a post-task questionnaire (see section C.2.3 and Appendix 5).

Furthermore, subjects who had used the screen under 500 lux (Group 5) had the additional task of setting the brightness and contrast control settings to their preferred levels as described in section C.2.5.

The results from the post-task questionnaire are shown in Table 19. The overall answers of Group 3 (from Experiment 2) and Group 5 (from Experiment 3) differed little, although many more subjects working under 300 lux (Group 3) reported having had aches in their shoulders or elsewhere.

There was no difference in subjective impressions of any screen characteristics when the screens were compared under the two lighting conditions (see Table 20).

The subjects were asked to set the brightness and contrast controls under both lighting conditions. In four successive tests the contrast control was set by the experimenter to positions 3, 4, 5 and 6. Subjects were then asked to set the brightness control to their preferred level. The results are listed in Table 21 for both lighting conditions. No significant differences were obtained for the brightness settings, either with different contrast settings or under different lighting levels.

Subjects were then asked to set both the brightness and contrast control settings to obtain their preferred levels for the two settings jointly. Table 22 shows the results. Again there was a broad agreement between subjects and there were no significant differences between the settings under the two lighting levels. Broadly, subjects preferred the brightness control between position 9.0 and 9.5 and the contrast control between positions 4.0 and 6.0.

To discover the brightness levels and contrast levels corresponding to the settings on the controls the experimenter took luminance measures on the screen using a Hagner photometer, when three large sections of the screen were illuminated at brightnesses respectively corresponding to the brightest character, the dullest characters and the background. Luminance measures were made using the Hagner photometer with a 1 degree angle of acceptance. The results are shown in Table 23, which shows the luminance and contrast ratios obtained in this way. The hatched area represents the subjects' preferred levels of both luminance and contrast. Figure 25 shows the relationship between the control settings and the luminances obtained.

It is apparent that subjects preferred the brightest characters to be between 300 and 525 cd/$m^2$, the dullest characters to be between 190 and 240 cd/$m^2$ and the background to be between 160 and 175 cd/$m^2$.

It must be stressed, however, that the ranges of values given above were obtained in a situation where there was an approximate 2:1 ratio between the brightnesses of the bright and dull characters. This ratio had been pre-set at this level by technical personnel in LME and at this level it was difficult to ensure that the dull characters could be seen at the same time that the bright characters were still clear. If the brightness was increased to enable the dullest characters to be seen, the bright characters became too bright and became blurred. Due to this fact, subjects were very restricted in the range of brightness and contrast levels which were acceptable. It was considered that if the ratio between the brightnesses of the bright and dull characters had been reduced (e.g. to approximately 1:1.5 or 1:1.25) it would have been possible to see both sets of characters clearly within a wider range of brightness settings than was possible in this experiment.

### C.5.3 Discussion

There were no significant differences between the task performances of the groups of subjects using the B1 screen (P39 phosphor plus matt green polaroid filter) under the two lighting levels of 300 and 500 lux. Neither were there any significant differences between the subjective judgements of the screen on the eight screen characteristics under the two lighting levels. The results of the post-task questionnaire indicated no major differences in eye fatigue between the two groups of subjects. More subjects who had used the screen under 300 lux (Group 3) reported pains in their shoulders or elsewhere, but this was unlikely to be attributable to the only experimental variable manipulated (the lighting level). It was more likely to be attributable to differences in age and acuity between the groups or between the frequency of spectacle wearing, as shown in the results from the pre-task questionnaire and from visual screening.

Thus it seemed reasonable to conclude that the datascreen (B1) could be operated equally effectively under 300 and 500 lux of ambient illumination. However, were older operators required to use the B1 screen (or indeed any other) for lengthy periods under 500 lux with the distressing illumination conditions used in Experiments 2 and 3 (i.e. light reflected over the subject's shoulders onto the screen), it was considered likely that their performance would suffer and that visual fatigue symptoms might become manifest.

The results of the test involving adjustments to the brightness and contrast controls indicate that, given the present 2:1 ratio between the brightest and dullest characters, a narrow range of control settings was acceptable.

| | |
|---|---|
| Brightness control | 9 - 9.5 |
| Contrast control | 4.0 - 6.0 |

Corresponding to screen luminance of

| | |
|---|---|
| Bright characters | 300 - 525 $cd/m^2$ |
| Dull characters | 190 - 240 $cd/m^2$ |
| Background | 160 - 175 $cd/m^2$ |

It was considered, however, that the pre-set ratio between the brightnesses of the brightest and the dullest characters should be reduced somewhat to improve legibility of both sets of characters and to increase the ranges of acceptable control settings. It was also considered worthwhile to provide an adjustment for the ratio between the brightness of the bright and dull characters. Such a control could ensure that during training the field labels could be set at a brightness only 10-20% lower than the brightness of the newly entered characters. As the operator learned the format of the ticket the brightness of the dull characters could be successively reduced.

## C.6 EXPERIMENT 4 : VERIFICATION OF SCREEN DESIGN ON PROTOTYPE WORKPLACE AND PRELIMINARY ESTIMATION OF VISUAL FATIGUE RESULTING FROM PERIODS OF SCREEN USE

Experiments 1 and 2 led to the selection of a datascreen for ANE 403, (Monitor with P39 Phosphor and a glossy screen fitted with a matt green circularly polarising filter). Experiment 3 showed that this datascreen could be used equally well under 300 or 500 lux of ambient illumination. However, the first three Experiments were carried out on a workstation which was not the design recommended for ANE 403 in Phase I. Also, the lighting in these first three experiments was arranged to ensure maximum screen reflections of light fittings, in order to sensitise comparisons between different datascreens and their treatments.

The purpose of Experiment 4 was to verify that the chosen datascreen would prove effective under better lighting conditions than those used hitherto and when used at the workstation recommended in Phase I. It was also desirable to require each subject to work for four twenty minute periods (rather than six five minute periods used in Experiments 2 and 3). It was felt that this longer exposure to the screen while performing a visually demanding task would provide a useful test of visual fatigue.

### C.6.1 Experimental Method

Essentially the methodology employed in Experiment 4 was similar to that used in Experiments 2 and 3. However, the following differences should be noted:-

1. The workstation used in Experiments 2 and 3 was replaced by an ANE 403 prototype, designed according to the recommendations provided in Phase I.
2. The lighting conditions were improved for Experiment 4 to provide better conditions, notably reducing reflections of the lights in the screen.
3. The task was carried out by the subjects for four periods of twenty minutes each rather than six periods of five minutes each.

Experimental design

Experiment 4 was simply an extension of Experiments 2 and 3. The datascreen which proved best in Experiment 2 and which was tested under 500 lux in Experiment 3, was re-tested in Experiment 4 with a new group of experimental subjects (Group 6) under different experimental conditions specified above. Since basically the same experimental methods were employed in Experiments 3 and 4 the results obtained in the two experiments were broadly comparable.

Subjects

The ten female subjects who took part in the experiment had not participated in Experiments 1, 2 or 3. Their judgements and subjective feelings of fatigue (but not their performances) were to be compared with those of the ten subjects who had been tested in Experiment 3. As in Experiments 2 and 3, data were collected on the subjects' eyesight and proneness to symptoms of visual fatigue, but, again these data were not available in time to use them to match subject groups. As in

Experiments 2 and 3 the data were used to make post hoc comparisons between the subjects in each group.

Apparatus

The datascreen used in this experiment was that used in Experiment 3, a monitor with a P39 phosphor and a glossy screen, over which was fitted a matt green circularly polarising "Polaroid" filter 1.83 mm thick (screen B1), bent to touch the screen as much as possible.

The datascreen was mounted upon a prototype workstation of the design recommended in Phase I.(See B.5). Essentially, it consisted of a single surface, adjustable between 580 and 720 mm, upon which was mounted the keyboard and a wedge-shaped plinth which held the front of the screen at a height of 180 mm above the workstation and enabled it to tilt backwards by up to 22$^{0}$ (see Figure 26).

The performance task which the subjects carried out using the datascreen was similar to that used in Experiments 2 and 3 (see Sections C.4.1 and C.5.1) and described in detail in Section C.2.4. The only difference was a procedural one. Subjects worked at the screen for four twenty-minute periods rather than six five-minute periods. The interval between each twenty-minute period was three minutes, during which time the results were printed on a hard-copy printer. The task material and data collection techniques were identical to those used in Experiments 2 and 3, (see Section C.2.4 for details).

The light level of 500 lux was produced in the same way as in Experiment 3 from six twin-tube fluorescent light fittings with 40W tubes (see Figure 24). The mirror used in Experiment 3 to produce strongly directional light over the subjects' shoulders in order to maximise reflections was omitted in Experiment 4.

As in Experiments 2 and 3 the following apparatus was also included:-

(a) a pre-task questionnaire (see section C.2.1 and Appendix 2).

(b) a Bausch and Lomb Orthorater (see section C.2.1 and Appendix 3).

(c) a series of rating scales to evaluate screen characteristics (see section C.2.2 and Appendix 4).

(d) a post-task questionnaire (see section C.2.3 and Appendix 5).

The purpose for which these various instruments were used are described in the sections referred to above.

Procedure

The experimental procedure was identical to that used in Experiment 3 except for two minor modifications.

1. There were four task periods of twenty-minutes duration each, rather than six five-minute periods.

2. The subjects were not asked to adjust the brightness and contrast controls.

As in Experiments 2 and 3 there were three experimental stages. In the first stage the company optician conducted a visual screening of each subject who then completed a pre-task questionnaire. During the second stage subjects performed a visually demanding task for four periods of twenty minutes each with a break of 3 min between each period. As in Experiment 2 and 3 subjects were informed of the number of errors they had made in each period. The third stage involved subjects completing two questionnaires. In one they recorded their subjective impressions of the screen on eight characteristics (see Section C.2.2 and Appendix 4) In the other questionnaire they reported their subjective impressions of visual fatigue etc. (see Section C.2.3 Appendix 5).

### C.6.2 Results

Pre-task tests: Subject differences

As discussed in Section C.6.1 each subject was given a visual screening test by the company optician and asked to compare the pre-task questionnaire before going on to the main experimental task. Unfortunately, as in Experiments 2 and 3, it proved impracticable to use the results of the pre-task questionnaire and the screening test to match the subject groups. The results of these pre-task tests were, however, used for post hoc comparisons of the subject groups.

The results from the visual screening test are summarised in Table 24. It will be apparent from this table that more subjects in Group 6, (who used the screen B1 under better illumination conditions of 500 lux) wore glasses, and more had far and near visual problems even when wearing their glasses. The results from the pre-task questionnaire (Table 25) indicate that the average age of subjects in Group 6 was higher than that of subjects in Group 5. It can also be seen from Table 25, that, in addition to there being more spectacle owners in Group 6, these owners had also worn their spectacles for a longer period of time. Both spectacle wearers in Group 5, however, reported that they could not wear their spectacles for long periods of time, possibly because these spectacles corrected for distance vision and were inappropriate for close work (see Table 24).

There were no major differences between the groups in the number of subjects reporting specific visual fatigue symptoms, headaches, or pains when reading. However, more subjects in Group 6 reported eyestrain when reading, watching t.v. and watching cinema; but more drivers in Group 5 reported eye-strain when driving. Large proportions of both groups reported that their eyes were irritated by lights but more of Group 6 reported that their eyes were irritated by cigarette smoke.

In summary it may be stated that more subjects in Group 6 had poor eyesight and wore spectacles as a result. Moreover, a greater number of subjects in Group 6 reported experiencing eyestrain in visually demanding tasks and they also reported greater irritability to cigarette smoke. Thus, it might be expected that this group would report more eyestrain after working at the datascreen for four periods of twenty-minutes each.

Visual task performance

The results of the performance of the subjects on the visual task were not collected for comparison with the performance of subjects in other experiments. The visual task was used in Experiment 4 to try to induce as much visual fatigue as possible using the chosen data screen in a task designed to match or exceed the demands of that involved in a semi-automatic telephone exchange.

The visual task also provided data on subject performance over four successive work periods. The results from the four work periods have been treated separately, to measure any changes in performance which might be brought about by fatigue.

Table 26a shows the average percentage error rates of the subjects over the four successive twenty-minute periods. The tendency for error rate to decrease over the four task periods is not significant, as shown in the Analysis of Variance Summary Table (Table 26b). Tables 27a-29a show the average response times for all responses (Table 27a) for correct responses (Table 28a), and for incorrect responses (Table 29a) over the four task periods. It is clear from the accompanying Analysis of Variance Summary Tables (Tables 27b-29b) that no significant differences in response time were detected over the four twenty minute periods.

Post-task tests: Visual fatigue and judgements of screen characteristics

The results from the post-task questionnaire (Table 30) showed no large differences between the subjects who had used the screens under the two conditions. This indicates that the lengthier experimental period in Experiment 4 (1 hour 20 min rather than 30 min as in Experiment 3) and the poorer eye condition of some subjects in Group 6 (see Tables 25 and 25) were at least compensated for by the less severe lighting conditions and the better designed workstation used in Experiment 4.

The results from the subjective ratings of the datascreen under the two conditions (Table 31) indicate that, as might be expected, the less severe lighting conditions led to very few reflections of lights being seen in the datascreen (significantly better at 0.01 level).

## C.6.3 Discussion

As in the previous experiments in this series,it was not possible to match the subjects in Experiment 4 with those in Experiment 3. In post hoc comparisons it became clear that subject Group 6, which used the B1 datascreen under the better lighting and workstation conditions of Experiment 4, was different in several respects from subject Group 5. Subject Group 6 had a higher average age than subject Group 5 and also contained a higher number of people with visual problems which were only partially corrected by spectacles.

The subjects in Experiment 4 (Group 6) worked at the datascreen for a total of 80 minutes, almost three times as long as Group 5 had worked at the screen in Experiment 3 (30 minutes). Since Group 6 contained more people with poor vision, had a higher average age and worked at the screen for much longer, one would expect that subjects in this group might report many more symptoms of visual and postural fatigue despite the better lighting conditions and workplace design. That such a difference in fatigue was not observed indicated that the screen could probably be used for periods of up to 1½ hours without any major visual or postural fatigue manifesting itself, provided short breaks were given.

In the discussion on Experiment 3 it had been considered that an older group of subjects using the datascreen B1 under 500 lux illumination might have performed less well and reported more visual fatigue. These conditions existed in Experiment 4 but no drop of performance or increase in reported visual fatigue occurred.

## C.7 DISCUSSIONS AND CONCLUSIONS

The purpose of this Section is to summarize the methods and results of the four experiments described in the preceding Sections and to state the final conclusions which may be drawn from them:-

The discussion will be concerned with three major aspects of the experiments:

1. Absolute levels of measurement.
2. Relative levels of measurement.
3. Experimental shortcomings and consequent reservations about the results.

These three aspects are dealt with separately in the next three sub-sections which are followed by a final sub-section containing the conclusions.

### C.7.1 Absolute levels of measurement

Before describing the relative differences between the datascreens and between the other experimental conditions it is necessary to examine the results with reference to criteria outside the experimental situation. This may be done conveniently in respect of three sets of information corresponding to the three phases of Experiments 2 - 4 inclusive.

1. Pre-task information (the condition of subjects' eyes before the experiment).
2. Task performance (the performance of subjects on the visual task).
3. Post-task information - The visual fatigue experienced by subjects after working at the datascreen for several periods.

It will be noted that the subjective judgements made by subjects on eight screen characteristics have been excluded from this part of the discussion. Because few of the subjects used in this experiment had any experience with datascreens and because subjective judgements are by their nature comparative it was felt that the discussion on them should be restricted to Section C.7.2 on relative levels of measurement.

<u>Pre-task information</u>

The results of the visual screening carried out by the Company optician (see Tables 5, 13 and 24) indicate that many of the subjects used in these experiments had eyesight which fell below the Purdue Standard for people doing Administrative and Clerical Jobs, even when they wore their spectacles or contact lenses. In fact, several individuals had seriously lowered visual acuity in one eye (see Numbers suffixed by 'R' in Tables 5, 13 and 24). Although, as will be stated later, there is no evidence that the subject groups bear any degree of similarity to telephone operators it is possible that this latter population may have similar visual problems which may lead to short or long term performance decrements or minor, but inconvenient, health risks when the individual is given a visually demanding job.

The results from the pre-task questionnaire given to all subjects before they carried out the visual task in Experiments 2 - 4 are shown in Tables 6, 14 and 25. It is apparent that several subjects had such a severe eye complaint that they were aware of it. The lack of colour blind subjects can be explained by the fact that the subjects were all female (colour blindness is relatively common in males but very uncommon in females).

The answers to question D1 (Tables 6, 14 and 25) give cause for most concern. High proportions of the subjects used in these experiments reported having at some time experienced visual fatigue (tiredness in eyes, eyestrain, etc.) or symptoms which may be associated with fatigue (soreness, itching, twitching, heaviness, blurring of print, etc.). Most subjects also reported getting headaches,although the majority only had them occasionally. Few subjects were willing to attribute their headaches to eyestrain but it is possible that telephone operators may blame headaches on the screen even if they could be attributable to other causes. Pains in the neck, shoulders and back when reading are reported but cannot be attributed solely to poor posture. Quite a number of subjects report eyestrain in visually demanding situations,such as reading, watching TV or the cinema, and driving. Again, the causes of fatigue (posture, stress, visual difficulties) cannot be separated, but it is sufficient to know that visual and postural fatigue are relatively commonly experienced in everyday life.

The number of subjects reporting that their eyes were irritated by lights and cigarette smoke indicated the need to provide good environmental conditions for operators of ANE 403 because it was possible that such symptoms would be in part attributed to the datascreen.

Task performance

The task used in Experiments 2 - 4 was not designed closely to simulate the telephone operators' task in ANE 403. The primary purpose of the task was to tie the subject visually to the screen for periods, in order to examine visual fatigue, if it occurred. The secondary purpose of the task was to provide an objective measure of relative legibility (see Section C.7.2). The task was made simple in order to try to minimise subject differences and to ensure that error rates were low, so that response times could be used to assess relative legibility. Surprisingly, error rates were quite high. Some subjects recorded error rates of nearly 20%. Average error rates of 2-6% for subject groups seem alarming, but may in part be due to the boring nature of the task, which may have encouraged subjects to risk making errors in order to work quickly. There appeared to be a trade-off between error rate and response time. The response times cannot be used as an absolute measure because they are highly determined by the nature of the task.

Post-task information

The post-task questionnaire results (Tables 11, 19 and 30) indicated that many subjects experienced symptoms of visual fatigue after using the screens for a time, and three subjects reported having a headache which arose during the task.

Almost all subjects reported some tiredness in their eyes (Question C1), and a number reported postural pains (Question B4). Most subjects felt that they would have liked to pause. Some, 30-50% depending on subject group, indicated that they would have liked to pause every ten or twenty minutes. Although the majority expressed the feelings that it was easy to work with the screen for the time they were involved (thirty or eighty minutes) more than 30% in five of the six groups stated that it was not easy to work with the screen for that period of time.

C.7.2 Relative levels of measurement

One of the primary purposes of the study was to select a datascreen from those available and to test its suitability on the basis of experimental evidence. Thus, the experiments reported provided a number of results which could be used to make comparative judgements (some of which could not be related to any absolute criteria in any very simple manner).

The performance of subjects on the visual task was very variable, and swamped any differences that might have existed between the screens. No significant differences in error rate or response times were found between the four screens tested in Experiment 2, or between the various experimental treatments applied to screen B1 in Experiments 3 and 4. This result may be explained, in part, by the selection process used to select the datascreens which only brought forward design solutions which were considered worthy of further investigation.

Although the visual fatigue of subjects was apparently high following periods of screen use (see Section C.7.1), no large differences were detectable between the fatigue reported by subject groups who had experienced the different experimental treatments.

The major differences between experimental conditions were detected in the subjective assessments made of important screen characteristics (such as flicker). In Experiment 1, these subjective judgements were the sole method for selecting four out of eight datascreens. In Experiment 2, datascreen B1 gained high ratings on five out of eight of the screen characteristics and consequently it was chosen for Experiments 3 and 4. In Experiment 3, the subjective judgements on the datascreen B1 at 500 lux ambient illumination were not significantly different from those obtained from another group of subjects viewing the same screen under 300 lux of ambient illumination. The improved lighting and the recommended workstation used in Experiment 4 led to a significant improvement in the subjective judgements of reflections of lights in the datascreen but no other significant changes in subjective judgements were apparent.

### C.7.3 Experimental Shortcomings, and Reservations about the Results

The experiments reported here were carried out as an interim study, to check that the datascreen design suggested in Phase I was potentially practicable and to provide ergonomics advice for the selection of a datascreen design which might be used in ANE 403. There were, however, several shortcomings in the present experimental methodology which were noted and which gave rise to some reservations about the results. It should be pointed out that these shortcomings were a result of practical difficulties at LME and the problems of conducting experiments at a remote location in another country where no appropriately qualified staff were available to assist.

The primary shortcoming was in a lack of time to plan, conduct and analyse individual experiments. With only four weeks to carry out four experiments, equipment was inadequately prepared, subjects were not properly selected or briefed fully, and decisions between one experiment and the next had to be based on very sketchy analysis.

Before the experiments began the availability of equipment was not certain, and consequently even the first experiment had to be hurriedly planned and executed with little time for a pilot run. Since the succeeding experiments were interdependent, it was necessary to make decisions on each experiment based on the results of the previous experiments. Little time was available between each of the four experiments to carry out full analyses of results and to reflect on them before specifying the conditions of the next experiment.

The treatments applied to the datascreens were particularly poor. The matt finish applied directly to the monitor with the P39 phosphor (screen BO) had been poorly prepared, and that, together with the fact that the monitor was badly adjusted, led to the early rejection of a datascreen which might have performed well in subsequent experiments. The polaroid filters and the 3M louvred filter were heavy gauge plastic sheets which would only bend in one plane, and consequently could not be fitted closely over the whole of the screens. Such a poor fit was particularly undesirable in the matt finished filter, since it is probable that it caused severe blurring of the characters in places where the gap between screen and filter was large.

The subjects were necessarily few, given the severe time limits on the experiments, and, unfortunately, sufficient information was not available in advance to match subjects in groups on the basis of age, eyesight, experience or other attributes considered to be important. Moreover, the subjects were not all telephonists and their characteristics bore an unknown relationship to this target population. In post hoc analyses, the shortcomings of the approach adopted (taking subjects as they came) was amply demonstrated. Significant differences in average age, in the frequency of eyesight problems, etc. were evident in the subject groups studied. This made it rather difficult to ascribe differences in performance or judgement solely to differences in experimental conditions, since the subjects who used them were not matched.

The task itself was a little boring and it was probably difficult for subjects to recognise when they were doing well or badly. The task was designed to be simple, in order to minimise the time needed to train the subjects and to reduce the variations between subjects to a minimum. Unfortunately, inter-subject variations still swamped any differences there might have been between experimental conditions. The boring nature of the task, rather than its primarily visual component, may have led to the feelings of fatigue and to preferences for frequent rest pauses.

## C.7.4 Conclusions

It was apparent from the results of the eye tests that many people in clerical jobs in Companies such as L.M.E. probably have moderate or even severe visual problems. It is considered likely that such problems may be magnified or may lead to further problems if such people are required to work at tasks involving lengthy use of datascreens. It is necessary to try to isolate the types of visual problem which are particularly important in tasks where datascreens are to be used continuously or very frequently.

The fairly high error rates obtained in the visual task used in Experiments 2-4 indicated the need to obtain comparative data on a similar task involving numbers presented by another medium, e.g. the printed page, to ascertain if the results were due to the task or to the datascreens. Unfortunately no facilities could be obtained for such a study.

The high degrees, in absolute terms, of visual fatigue reported by subjects after working at the datascreens for periods as short as thirty minutes were cause for concern. It would have been interesting had the subjects who performed the task outlined in the previous paragraph been similarly asked to report fatigue immediately the task was completed.

Experiment 1 led to the selection of four datascreens which were then tested further in Experiment 2. As a result of Experiment 2, Screen B1 (P39 Phosphor screen fitted with a matt green circularly polarising filter) was selected from the four tested, primarily on the basis of significantly better subjective ratings. In Experiment 2, the datascreens were tested under very severe lighting conditions of 300 lux. Screen B1 was retested with a new group of ten subjects in Experiment 3 under severe lighting conditions of 500 lux, and no significant changes were found in task performance, subjective assessments or visual fatigue. The screen was then tested with a new group of subjects, under better lighting conditions and with a better workplace design, but for longer periods of time. No significant changes in task performance were detected over four twenty-minute periods. Visual fatigue after eighty minutes of screen use under better conditions (experiment 4) was not significantly worse than after thirty minutes under poor conditions. (Experiments 2 and 3). Thus it was apparent that, although the absolute levels of task performance and visual fatigue gave rise to some concern, the datascreen B1 seemed to give similar levels of performance and fatigue under quite variable lighting and other conditions. Thus, it was recommended that, consequent on the resolution of the technical problems, indicated below, screen B1 be used for ANE 403.

One of the datascreens (screen BO) was rejected as a result of Experiment 1. However, this was a poor sample with a poorly applied matt coating and badly adjusted electronics (which led to spot wobble). Since it was believed that the polaroid filter might subsequently be rejected on practical grounds (i.e. that it could not be fitted closely to the datascreen) it was pointed out that screen BO should still be considered. Screen BO, it will be remembered, had a P39 Phosphor with a matt finish applied directly to the screen surface and on a number of technical and ergonomic grounds was considered to be an effective design. It was recommended that a better sample of screen BO should be tested under near identical conditions to those used in Experiments 2, 3 and 4. In the event this screen was chosen for ANE 403 without any further tests.

It was pointed out that, whichever screen was selected, it would perform best in a well designed lighting environment with the direction of the light restricted to a primarily downward direction, preferably within a solid angle of $120^{o}$ vertically downward.

The other important environmental conditions, temperature, humidity and air circulation, were considered to be equally important in reducing the degree of visual fatigue experienced by datascreen users.

It was recommended strongly that, whichever datascreen was selected, it should be more fully tested in a task more closely resembling that of the telephone operator. It was considered essential that the task should be performed for periods of up to two hours, that objective and subjective measures of visual fatigue should be used as well as performance measures, and that these measures should be monitored throughout the experimental periods.

# CHAPTER D - PHASE III

## PROTOTYPE TESTING USING SIMULATED CALLS

### D.1. PREFACE

Phases I and II of the study were concerned with arriving at detailed ergonomics specifications for the design of the ANE 403 operator position. During the two years following completion of Phase II, Company engineers put most of the recommendations into practice, requesting further help from the author whenever design conflicts arose or details required additional clarification.

One problem which arose as a result of Phase II was that the datascreen design strongly recommended as a result of this phase had to be rejected on technical grounds. It will be seen in Section C.7.4 that, as a result of Phase II, it was recommended that Screen B1 was adopted for ANE 403, consequent upon the resolution of a possible technical problem. Screen B1 was a datascreen with a tube coated with P39 phosphor and fitted with a matt green circularly polarising filter (Polaroid Corporation - reference No. HGCP 21). However, this filter was 1.83mm thick and thus could not be bent to touch the datascreen at all points. Where the filter was not touching the datascreen it tended to cause blurring of the characters, a characteristic deemed unacceptable for the ANE 403 system. It was found impossible to mould, or otherwise to fit, the polarising filter to the datascreen, and thus datascreen design B1 had to be rejected. The datascreen suggested as a possible alternative to B1, in the event that the technical problem could not be resolved, was datascreen design BO. This consisted of a P39 coated tube, with a matt finish applied directly to the curved tube face. This alternative had been eliminated in Experiment 2 of Phase II, but it was known that the sample used in Phase II had been poorly prepared and that the matt finish was particularly inadequate on the sample used. Improved samples of design BO were forthcoming from the manufacturers, and LME considered that no further experimentation was required to ascertain the adequacy of this design for ANE 403.

From the beginning of the project it had always been stressed that some of the ergonomics recommendations could only be made on limited evidence and that some design conflicts required untested compromise for their solution. Thus, it had been recommended in the initial stages of the project that a fully functioning prototype of the operator position should be tested under realistic call handling conditions and for lengthy periods of time (see Section A.8). It was impractical to build a system which could be interfaced with an existing exchange, which would have allowed the system to be tested with real telephone calls. Another alternative method which was considered was to use actors to play the parts of the calling and called subscribers. However, this arrangement would have required at least two actors per call-handling subject for the entire length of the experiment and would also have required complex organisation and the provision of several communication channels. However, by the Autumn of 1979 four prototype operator positions (terminals, desks, chairs, etc.) were available, housed in a pleasantly decorated, air conditioned room designed to conform to the lighting conditions specified in Phases I and II. These operator positions were linked to a computer, to a magnetic recording/playback device and to a supervisor's position which was also housed in the experimental room. The computer was programmed to provide an operator training system which presented datascreen information and which responded appropriately to operator commands via the keyboard. The computer also fed pre-recorded subscribers voices and exchange tones to the headsets of the operator at theappropriate moments during call handling. Moreover, details of the performance of each subject could also be recorded by the computer and could be displayed at a supervisor's desk when requested. At the suggestion of the author this facility was constructed as a dual purpose device. It was intended to be used as a simulator in those experiments and later as a training device for operators. Thus, the facility provided the ideal opportunity to test the prototype system under reasonably realistic conditions within practical limits of time and cost.

Phase III of the project involved groups of subjects handling a variety of telephone calls types for substantial periods. All the subjects used were telephone operators, either from LME or from Televerket, the Swedish telephone administration. The experiment comprising Phase III enabled these operators to work with the ANE 403 operating position and thereby to assess its merits and to comment upon difficulties encountered. This Phase also enabled the experimenter to collect subjective ratings of visual fatigue before and after the experimental periods, and to compare these with ratings obtained at similar times during the operators' normal day working with 'cordless' exchanges.

## D.2. METHODOLOGY

A single experiment was conducted in this study and, to prevent confusion in subsequent reporting, it has been called Experiment 5, since it followed the four experiments already conducted in Phase II.

### D.2.1 Experimental Plan

Experiment 5 enabled 28 subjects to be observed over a period of two hours, handling a variety of types of call. The primary purpose of this experiment was to use a large subject group, critically to evaluate the equipment and task design of the operator position. This was done in terms of:

(a) Call handling rates and errors

(b) Adequacy of datascreen and keyboard

(c) Adequacy of workstation design

(d) Difficulties encountered in handling calls

(e) Degree of visual fatigue subjectively experienced by subjects

(f) The extent of visual fatigue as measured by the Mallet fixation disparity test administered before and after the experiment.

To maximise the number of subjects taking part in the experiment, while using the training system for the time for which it was available, it was necessary for each subject to be tested for a relatively short period of time, (i.e. two hours) after they had been trained. Twenty-eight subjects took part in the experiment after undergoing an ophthalmic screening test which enabled them to be placed in one of four experimental groups, within each of which subjects had similar eyesight.

### D.2.2 Subjects

#### General

All the switchboard operators at L.M. Ericsson and a number at Televerket, the Swedish Telephone Administration, were asked if they would like to participate in the experiment. A total of 31 subjects were eventually visually screened and the three who were omitted from Experiment 5 were omitted simply because they were unavailable during the period of the experiment. Of the 28 who took part in Experiment 5, 16 were from the L.M. Ericsson exchange and 12 from Televerket.

#### Experiment 5

As a result of this visual screening test, each of the 28 subjects who took part in Experiment 5 was assigned to one of four subject groups (denoted A-D).

| Group | Eyesight | Description | Type of Spectacles Typically Worn | No. |
|---|---|---|---|---|
| A | 'Normal' | Good acuity without spectacles | None | 6 |
| B | Hypermetropes | Corrected 'long' sight | Reading (Convex) | 9 |
| C | Myopes | Corrected 'short' sight | Distance (Concave) | 5 |
| D | Presbyopes | Corrected 'long' sight with poor accommodation at 'near' | Bifocal (Multi-lens) | 8 |

The results of the visual screening tests for each subject in these experimental groups is shown in Table 32. It was hoped that it would be possible to match these subjects on age, sex and occupation and to obtain equal numbers in each group. All the subjects were female telephone operators from the private exchange at L. M. Ericsson or from Televerket, the Swedish Telephone Administration. However, since some eyesight problems are age related, it was not possible to balance the subject groups on age. Unequal sized subject groups resulted from an insufficiently large pool of volunteer operators. However, it is possible that, because of the age related eyesight problems, even a large subject pool might not have enabled the subject groups to be matched by age.

Before beginning Experiment 5,all 28 subjects were asked to complete Questionnaire 3,relating to the general condition of their eyes (see Appendix 8). The overall results of this questionnaire are indicated by the frequency scores typed in italic numbers beside each of the answer boxes in Appendix 8. The frequency of the replies to Questionnaire 3 within each of the four subjects groups are summarised in Table 33.

### D.2.3 Apparatus

Operator positions and experimental room

The primary purpose of the study was to evaluate the performance and acceptability of the entire operator position. Thus,it was essential that the design of the equipment, the call handling dialogue and the environment conformed as closely as possible to the recommendations made in Phase I, taking into account subsequent developments impinging on the design as described in Phase II

The datascreen, keyboard and console conformed closely with the specification drawn up in Phase I. Unfortunately, the anti-reflective treatment applied to the datascreens before Experiment 5 was inadvertently cleaned off the datascreens by the cleaning staff at some indeterminate time during the Experiment.

The environment of the operator room conformed closely with the recommendations made in Phase I, except that the general illumination was raised to approximately 500 lux in line with the findings of Phase II. The four workstations were arranged in the form of a cross around a central plant arrangement enabling each subject to see each of the others without undue visual distraction (see Fig.27). The general temperature of the room was acceptable, but air movements were rather slow, since the experimental room was not fully air-conditioned.

<u>Rest room</u>

A rest room, made available to all subjects, housed a coffee machine and was close both to a washroom and to the test room.

<u>Training system</u>

It was essential that the task set to subjects in these experiments had to be as accurate a simulation as possible of the task of an ANE 403 operator. It was important that the form, content and timing of information received and transmitted by the operator were as realistic as possible. In the original proposals for this study it was thought necessary to use additional people as A and B subscribers, but this solution, while it might have brought substantial benefits in improving the quality of the simulation, was impracticable. In the event, the ANE 403 Training System formed the basis of the simulator and the A and B subscribers were simulated by recorded messages, whose onset and duration were controlled by the call sequence program resident in the control processor.

The ANE 403 Training System was used to produce the ticket format on the datascreen, to produce the appropriate field contents during call making, to initiate and stop audio messages, and to detect the actions of the operator through the keyboard. This system, designed for training ANE 403 operators, was independent of, and separate from, the ANE 403 system as incorporated into exchanges. The following paragraphs explain briefly how the training system operated.

The ANE 403 Training System enabled simulated traffic cases to be prepared in advance, and to be fed to an operator at any of up to 14 normal operator positions, either individually or in sequences prepared by the teacher. Once calls were presented to the trainee they could be practiced with or without assistance from the teacher, who could be contacted by the subject by entering a request in the INQUIRY field. The teacher could monitor the call handling of any individual trainee in terms of datascreen, keyboard and acoustic messages and was able to collect simple statistics on the call handling performance of all the trainees over all sequences involving up to ten types of call at any one time (see Figure 33 for a sample printout of these statistics).

The traffic cases were prepared in advance on a special form and the information was transferred to a magnetic tape. From this tape the information was fed into the control processor. Up to 16 separate audio messages could be recorded on a special purpose magnetic disk from which they could be fed to the operators' headsets on instruction from the control processor during the call sequence.

The execution of the traffic cases was handled by the control processor which worked completely according to the prepared data. In those data were specified what the control processor should write on the trainee's datascreen, which audio messages should be routed to the trainee and for how long, what the trainee should write onto the datascreen and the sequence in which function keys should be operated.

When a trainee was at a training position and had plugged in the headset, a message on the datascreen requested her signature. Once the signature has been entered and the FREE button pressed, the appropriate traffic cases were fed to that position in the sequence prepared by the teacher. The first call was initiated at the trainee position after a predetermined interval programmed by the teacher. The first call appeared in ticket form and, in the present situation (the most likely procedure to be widely adopted), acoustic connection was not established until the operator pressed the SM (Special Monitor) key. The execution of the traffic cases could be seen as a form of dialogue between the trainee position and the control processor. The trainee performed the different handling procedures step-by-step, where every step was checked by the control processor. Steps were acknowledged by a change in the information on the datascreen and/or by an acoustic message through the trainee's headset. A correct step was acknowledged in the same way as if it were a real traffic case, while incorrect steps were indicated by a brief ERROR message at the bottom of the display. When the call had been successfully handled, and the FREE button was finally pressed, the next call in a pre-arranged sequence was fed to that position after a pre-ordained time period.

The rigidity of the pre-ordained call handling sequence meant that the teacher had to determine the single 'correct' call handling procedure which, when deviated from, led to an error being recorded and an error message being displayed. In the real ANE403 system the operator is able to deal with calls in any of a variety of ways, all of which can be justifiably argued correct. Thus, some of the errors recorded in the experiments were artificially introduced by the training system. However, since most administrations are likely to insist on a fairly rigid call handling procedure, this criticism of the training system is not well justified.

Only ten call types could be resident in the control processor at any time but these could be presented in any sequence, with each call being repeated up to ten times within that sequence. The call sequence itself could also be repeated up to ten times. Each call type had an alternative procedure written into the program and this alternative could be selected at any time on command from the teacher position.

Call types

Ten simple call types were prepared for Experiment 5, each with an alternative ending. Thus, altogether, the teacher prepared twenty different calls, together with the necessary acoustic tones and messages.

The ten original and ten alternative call types used in Experiment 5 are described in Appendix 9a. They included demand calls where B answered, was busy or didn't answer and several demand calls requiring special services (e.g. personal call). They also included several types of call booking and search for booked calls. Finally, they included a number of delay calls.

In order to simplify the switch from one set of calls to the alternative set, this operation was done simultaneously for all ten call types during the ten-minute break between successive one-hour sessions. This was done because practical difficulties in the computer program prevented the selection of individual call alternatives while they were being handled by any subject.

Questionnaires

Three basic questionnaires were prepared for this experiment. These have been numbered 3-5 to distinguish them from the questionnaires used in Phase II. The first questionnaire (Questionnaire 3) was administered to all subjects before they began the experiment. This questionnaire inquired into the general eye condition of the subjects, asking for details of their spectacle or contact lens prescriptions (if any), for information on the extent to which they suffered various symptoms of eye fatigue, headaches, etc. Appendix 8 shows the complete English translation of Questionnaire 3 together with the frequencies of answers given across all the subjects.

Questionnaire 4 enquired about the subjective eye condition visual fatigue and headaches, and postural fatigue. It was prepared in two forms, one for administration at the beginning and the other at the end of each experimental period. An English translation of Questionnaire 4 is shown in its complete form in Appendix 10.

Questionnaire 5 asked subjects to rate the operator position equipment, workstation, environment and procedures and to compare the ANE 403 operators' job relative to the subjects' normal operator job. This questionnaire was administered to each subject at the end of Experiment 5. An English translation of this questionnaire is shown in Appendix 11.

## Keystone vision screening apparatus

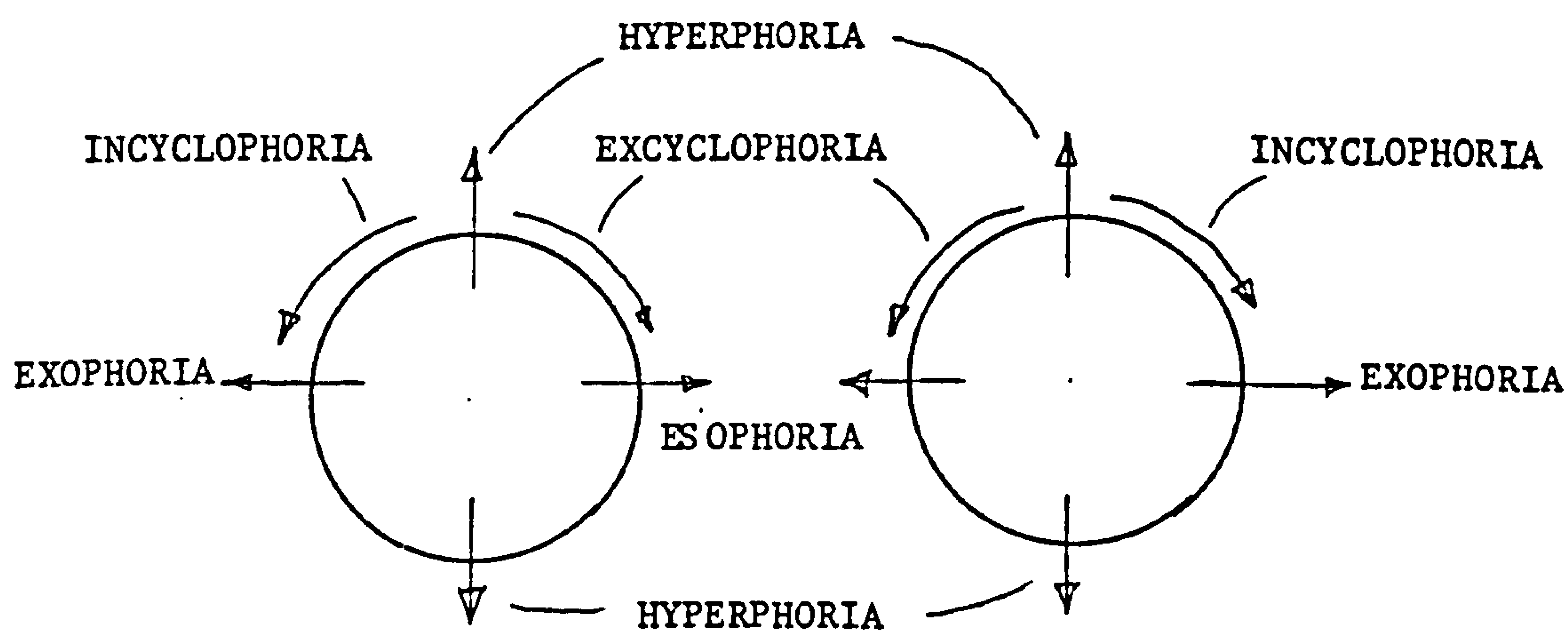
The ophthalmic optician used the Keystone Ophthalmic Telebinocular to conduct the vision screening of the subjects. Basically the unit shown in Fig.28, consisted of a forehead rest, binocular eyepieces, target card holder and illuminating lamp mounted on an adjustable base to cater for variations in the eye height of subjects. The targets consisted on photographically precise stereoscopic slides, which, when mounted in the unit presented a different 'picture' to each eye. The target holder of the unit used in these tests could be adjusted to provide a 500 mm viewing distance.

The 15-target set used for this screening test was the Visual Skills Test Set, widely used by eye specialists to determine the basic visual patterns before refraction (Figure 28). The targets provided accurate data on phorias, fusion readiness, binocular visual efficiency at far and near, stereopsis and colour vision.

The tester recorded the subjects responses to the various tests on a single form which provided a visual skills profile (Fig. 29).

## Mallett fixation disparity unit

Heterophoria is the inability of the two eyes to look towards the same point when they are dissociated (i.e. unable to form a common image). Heterophoria can be convergent (esophoria) or divergent (exophoria) in the horizontal plane, upwards or downwards in the vertical plane, inward rotating (excyclophoria) or outward rotating (incyclophoria) about the axis of vision. (These rotational phorias are sometimes termed tortional phorias). These heterophorias can be uniocular or biocular (see below).




In the last 20 years there has been a marked change in the clinical investigation of heterophoria since it has been recognised that it is equally important to determine to what extent the heterophoria is compensated by the opto-motor reflexes. For instance, it is not unusual for a person with only a prism dioptre or two of heterophoria to be plagued with headaches because their fusional processes are inadequate for even this low degree of imbalance; conversely, someone with 20 or more prism dioptres of heterophoria may be quite comfortable under the most adverse conditions (Mallett, 1964). Since the amount of heterophoria measured with dissociating devices may bear no relationship whatever to a person's symptoms, several ophthalmologists have devised means of examining uncompensated heterophorias in terms of 'retinal slip' or 'fixation disparity'. Fixation disparity, first recognised by Lau (1921), can be regarded as a minute strabismus but with the subject retaining fusion and stereopsis. Fixation disparity can be convergent, divergent, vertical or torsional. Of the two types of fixation disparity the most common variety occurs whenever a heterophoria places a burden upon the individual which the available fusional processes find difficulty in meeting (i.e. an uncompensated heterophoria).

As stated earlier the degree of heterophoria responsible for this situation may be of a low or high degree. Suppression of one eye, if it is present, is not particularly deep and will only appear sporadically. The disparity will always be small (i.e. it will be about 5 minutes of arc) and it may be present in one or both eyes. Mallett (1964) described his technique for detecting and identifying any fixation disparity. The Mallett fixation disparity unit consisted of a light-box which could be hand-held by a subject at their normal reading distance or placed at a fixed viewing distance. A retractable rule, fitted with the unit, enabled the viewing distance to be ascertained easily. As shown in Fig. 30, the front face of the unit contained an opaque white sheet printed with black text in a book-style font. Two round black targets were included in this text, each of which would subtend 10-30 minutes of arc at the operators' eyes when viewed at 200-500 mm (the recommended range of viewing distance for this apparatus). Each of the targets contained a rectangular white panel printed with the black letters OXO. In one target the panel and letters were horizontal and in the other they were vertical (see Fig. 30). Two green polarised strips were placed on either side of the white panels on each target, one above and one below the cross for detecting lateral disparities, and one to each side of the cross for detecting vertical disparities. In each target one of these strips was polarised vertically and the other was polarised horizontally, so that, viewed through a polarised visor with vertical and horizontal polarised filters in front of the subjects' eyes, each of the green strips of a target was only seen with one eye. Green was chosen for these strips because it more nearly conformed with Ivanoff's (1952) work on chromatic selectivity of the eye at the near point, and could not easily be suppressed. The printed panel was opaque and thus was illuminated by ambient light but the green polarised strips and the white OXO panel were translucent and were illuminated by lamps inside the unit.

The surrounding printed panel provided an appropriate background and helped 'fusion lock'. The central white panel and the printed OXO acted as a fusion lock since they could be seen by both eyes. However, since each of the two polarised strips on either target could only be seen by one eye, their non-appearance would indicate central suppression of the appropriate eye, while deviation in the position of any strip would indicate fixation disparity.

As stated above, the unit had to be used in conjunction with a polarised visor. Since this reduced the effective illumination somewhat, it was necessary to increase the ambient illumination by using a table lamp.

The fixation disparity test was performed with the subject wearing (if necessary) spectacles appropriate to a viewing distance of 500 mm (i.e. those spectacles used during the experiment). The unit was placed 500 mm from the subjects' eyes (as measured with the retractable tape incorporated into the housing of the unit) and the polarised visor was placed over their eyes. The subject was told to concentrate upon the cross in the middle of the word OXO in the illuminated left hand target and asked if he could simultaneously see two green strips, one above and one below the cross (remember, one strip was seen by the right eye and one by the left eye).

If only one green strip was seen,this indicated very dense central supression in the appropriate eye, (see Figures 30i and 30j In the great majority of cases, both strips were seen and the subject was then asked if the top green strip appeared exactly above the bottom green strip (Figure 30c). However, if the top strip (seen by the left eye) was to the right of the bottom strip, an exophoric disparity was present. In such cases the subject was asked if only one strip had moved from the cross (as in Figure 30g) demonstrating a retinal slip in one eye, or whether both had moved (as in Figure 30h) showing a slip in both eyes. If the slip or disparity was in the opposite direction the subject was shown to have uncompensated esophoria, (Figures 30k and 301).

The subjects' attention was then directed to the second target which was then illuminated. Again, while fixating the cross, the subject was asked if he saw two green strips, and, if so, whether they appeared exactly level (Figure 30d), showing vertical orthophoria or compensated hyperphoria. If one line appeared slightly higher than the other (the word "slightly" was emphasised), this indicated an uncompensated phoria. As the left-hand strip was seen by the right eye, it would appear higher than the right strip in uncompensated left hyperphoria. Again, it was noted whether this slip occurred to the right or to the left eye or was bilateral (Figures 30m and 30n).

Although cyclophoria is very prevalent in near vision, it seldom causes trouble (Mallett, 1964). Cases of uncompensated cyclophoria could be detected by the sloping of one, or both, green strips. Either test target could be used for this purpose. Figures 30e and 30f show two such possibilities using the target with vertically polarised strips.

Since any slip would only subtend 5-15 minutes of arc, it was important to stress that <u>any</u> displacement, however small it appeared, was important.

It would have been possible to measure the extent of any slip using compensating prisms but this would have entailed a lengthier procedure, and neither prism lenses nor prism bars could be obtained for the experiment.

D.2.4 Procedures

Eyesight screening

A subject panel of 31 people were given a visual screening test by an ophthalmic optician using the Keystone apparatus, (of these 31 subjects, 18 worked in the LME exchange and 13 worked for Televerket, the Swedish telephone administration). All 31 subjects in this panel were invited to take part in the experiment, but only 28 were available. As a result of the screening tests, these 28 subjects were divided into four unequal sized groups according to their eyesight and the types of spectacles they wore. In addition to measuring the visual acuity of subjects, the optician also tested the muscle balance of the subjects' eyes and noted any hyperphorias. The optician also subjected each subject to a Mallett fixation disparity test and noted which subjects had uncompensated hyperphorias, (see Fig.31 for translation of the form).

Pre-experimental design

Subjects arrived for experiment in random order and were randomly assigned to the four training workstations available.

When first arriving at the experimental room the subjects were briefed on the nature and purpose of the forthcoming experiment . The subjects then were asked to adjust their workstation according to a few simple ergonomics guidelines, explained verbally and illustrated graphically (Fig.32). The posture of each subject was then checked by the experimenter and workstation adjustments made where necessary. All the physical workstation adjustments were explained, including the ability to alter the brightness and contrast of the datascreens. Subjects were next asked to complete Questionnaire 3, which requested information on the general condition of their eyes and Questionnaire 4, asking what degree of visual fatigue they experienced in everyday life.

Training

The teacher then gave a complete description of the call handling procedures and the types of call which the subjects would have to handle. The subjects were then asked to sit around one operator position while the teacher described the ticket format and the keyboard before working through two calls. Subjects then went back to their desks and were again led through several simple calls.

Subjects were given a set of sheets including complete protocols for handling each type of call, country letter codes and 'order' codes. The teacher then prepared a sequence involving the call types which was then transmitted to each operators' position. Subjects worked through this sequence at least once during a period of two hours, with the experimenter and teacher available for advice.

Testing

Subjects were tested in two-hour sessions, each consisting of two one-hour periods separated by a ten-minute break away from the operator position. This test session was given on the afternoon of the day in which the subject was trained.

In this two-hour session, the subjects handled twenty types of calls, ten basic call types and ten alternatives. The ten basic calls were given in a random sequence repeated as often as necessary during the first hour. During the ten-minute break given to subjects the teacher programmed the computer to provide the ten alternatives, again in a random sequence which was given to the subjects during their second one-hour period.

During each one-hour period over the two-hour session, the teacher requested the call handling statistics to be printed out at approximately twenty-minute intervals. The statistics showed the number of calls handled in each twenty-minute period, the average call handling time for each type of call and the number of errors made on each type of call. (Fig.33 shows a sample printout from the computer for four subjects over a twenty-minute period.)

## Post-test period

At the completion of the two-hour test period in Experiment 5 the subject was asked to complete the visual fatigue questionnaire (Questionnaire 4) and the equipment and procedure assessment questionnaire (Questionnaire 5). Subjects were then again tested using the Mallett fixation disparity unit.

## D.3 RESULTS AND DISCUSSION

### D.3.1 Visual screening tests and subject grouping

Table 32 shows the results of the visual screening tests applied to all subjects before Experiment 5 and on the basis of which subjects were placed into four groups (A-D) according to their eyesight. To enable easy reference, Table 32 also includes details of the type of spectacles worn by each subject during Experiment 5 and the results of the Mallett fixation disparity test applied before and after Experiment 5. The results of the post-task Mallett test will be discussed later under the heading "Visual fatigue" (Section D.3.3).

It is clear from Table 32 that subject groups B, C and D each contained approximately equal numbers of subjects from the L. M. Ericsson switchboard (LME) and from Televerket (TV). The unequal numbers of subjects obtained from these two sources is only apparent in subject group A which contains five LME operators and one Televerket operator, but even this large level of inequality is not significant (by a Fisher exact probability test applied on a 2 x 2 matrix formed by LME/TV and subject groups A + B and C + D). The average ages of the four groups varied greatly and these differences were statistically significant at $p < 0.001$ ($X^2_r = 17.1, df = 3$). Many subjects in Group B and D were considerably older than subjects in the other two groups, reflecting the older peoples' lack of accommodation at near, and their consequent need to use a reading prescription for close work. Group D were presbyopes, who not only suffered from lack of near accommodation, but also had poor distance vision, and consequently had to wear either reading and distance spectacles as required, or multifocal lenses, usually bifocals.

The column headed type of eyesight indicates that all the subjects placed in Group A had reasonable eyesight without the need for correction, those in Group B were hypermetropes requiring convex lenses for correcting near vision for tasks such as reading, Group C contained subjects who were myopic,requiring spectacles for distance vision or for all tasks and Group D contained presbyopes who required two types of correction, for near and for far vision.

The next column indicates the type of spectacles worn during Experiment 5. The datascreen was viewed at a distance of 500-600 mm,which is considerably more than normal reading distances of 300-400 mm, and yet cannot be considered distance vision. Thus, subjects in the three spectacle wearing groups B, C and D, had a choice of whether to wear spectacles or not and, if they had several pairs (frequent in Group D), which pair they wore. This choice was left to each subject, since only they could be expected to decide what felt most comfortable. The subjects in Group A possessed no spectacles. Subjects in Group B and C could decide to wear their reading or distance spectacles as appropriate,and one subject in Group C chose to wear her soft contact lenses. In Group D, subjects wore either bifocal spectacles or reading spectacles, none chose distance spectacles even when they were available.

The three columns of Table 32 ranged below the heading "Type and extent of Heterophorias" indicates the presence of poor muscular balance as measured by the optician when the subjects two eyes were dissociated (i.e. without any fusional lock).

The columns headed "Type of fixation disparity before" and "Type of fixation disparity after" contains the results of the Mallett fixation disparity test before and after experiments.

Table 33 shows the results obtained from Questionnaire 3, the eye condition questionnaire given to all the subjects before Experiment 5. In this questionnaire subjects indicated if they suffered from any eye problems, whether they could see with both eyes and if they were colour blind. A larger proportion of subjects in Groups C and D reported eye problems than subjects in Group A and B, but this difference was not significant at the 0.05 level of confidence (Fisher exact probability test for Yes and No answers to question B1, for the summed totals of groups A and B, and groups C and D). Only one subject reported vision restricted to one eye and it was apparent from the ophthalmic tests that this was due to occasional suppression of one eye. No subjects were colour blind, a predictable finding for a female sample of this size. The questions in Section C related to spectacles and contact lenses, and so were not applicable to Group A who, by definition, wore no spectacles. Group C had worn their present prescriptions for a shorter time and had been to the optician more recently than Groups B and D. This finding may be primarily related to the relative youth of Group C, whose eyesight would be expected to be changing more quickly, and which could be expected to contain more subjects who had recently found the need to wear spectacles. Subjects in Groups B and C only had one type of spectacle (although one subject in Group C also had contact lenses), whereas the majority of subjects in Group D had at least two pairs of spectacles (usually a pair of bifocals for general use and a pair of 'reading' spectacles for close work). A larger number of subjects wore tinted spectacles than non-tinted spectacles, but it is not known if this is for ophthalmic or fashion reasons. Most subjects reported that they could wear their spectacles for long periods of time (C6a), and few reported fatigue symptoms as a result of wearing their spectacles (C7a). The answers to question D1 showed significant differences between the groups in the frequency with which subjects in these groups reported that they never suffered from the twelve symptoms listed ($p \leqslant 0.05$, Friedman 2 way analysis of variance $X^2 = 9.3$, df = 3). Higher proportions of subjects in Group D reported never suffering from symptoms and the rank order of the groups in descending order of reporting never having the symptoms was D,C,B,A. Thus, the group with the normal uncorrected sight, which was also one of the younger groups, more frequently reported having these symptoms seldom or often. Moreover, there were consistent differences in the frequencies with which the symptoms were reported; with no subjects reporting that they never suffered from 'Tiredness in the eyes', for example (this finding was significant at p 0.01 by the Friedman 2-way analysis of variance $X^2 = 29.0$, df = 11). Symptoms such as blurring of written text, twitching of the eyelids and watering of the eyes were relatively frequent symptoms, whereas double vision, excessive blinking and dry eyes were very infrequent in all groups.

Although there are differences between the groups in the frequency with which subjects reported having headaches (D2) these differences could not be tested for significance. The position of the headaches reported did not differ significantly between subject groups. The position of general pains when reading did not differ significantly, but it can be noted that more subjects in Group D, those wearing bifocals, reported postural pains when reading. There were also no significant differences between the groups in the proportion of individuals reporting pains when reading. There were also no significant differences in the reported incidence of eyestrain when reading, looking at t.v., visiting the cinema or driving. Few people in any group suffered from undue eyestrain in other conditions but large proportions of all groups reported discomfort from strong light. Few people reported always suffering discomfort from flickering lights or television, but many experienced this discomfort 'seldom' rather than 'never'. In most groups cigarette smoke in a room was only 'moderately' or 'a little' irritating, but only three subjects reported that it was never irritating. Few subjects responded that their eyes felt "especially dry in a room with dry air or when it is particularly dry outside".

D.3.2 Call handling performance

During Experiment 5 call handling performances were measured in terms of total call handling time and average error rate per call. Call handling times were measured in seconds, from the first appearance of the ticket to the successful 'freeing of the position to accept another call. Call handling errors resulted from incorrect field entries on the ticket or from key presses out of the pre-determined 'correct' sequence.

The time and error scores were recorded for each subject, handling each type of call during each twenty minute call handling period. This allowed analyses to be made on variations in performance between experimental periods (see below) and between call types (see next page). Since, as will become apparent, large subject differences were observed, it was decided to carry out further analyses to determine whether any measured subject differences correlated with these performance differences and therefore analyses were conducted first on subjects grouped according to eyesight (see later in this section) and second on subject age versus performance (see later in this section). The final part of this section covers an analysis of the correlation between time and error performance measures.

Experimental periods

During Experiment 5 each subject handled calls for six twenty-minute periods. Periods 1-3 took place during the first hour and periods 4-6 followed during a second hour after a break of ten-minutes. During each of the hour long sessions the experimental periods were only defined by the intervals between the times when call handling statistics were printed out on a remote hard-copy printer.

Ignoring any differences between call types (which will be examined below), and averaging call handling times over all 28 subjects, gave the results summarised in Table 34 and illustrated in Figure 34. It is apparent that call handling times tended to decrease over the six periods while their dispersion diminished. This indicates that subjects were still learning the task throughout the experiment but shows no evidence of substantial fatigue in terms of performance, either during each hour session or over the two hours of the experiment.

A summary of the average error rates for all subjects shows a similar picture (Table 35 and Figure 35) except that a substantial reduction of errors after Periods 1 and

2 was followed by four periods with no substantial change and no marked trend. It is thought that the majority of these errors were due to idiosyncracies of the training system. (see Section D.2.3).

Call types

In Experiment 5 subjects handled twenty different calls in two sets of ten, with each set presented exclusively during one of the two one-hour experimental sessions. Thus, during periods 1, 2 and 3, subjects handled ten types of call. During the ten-minute break between sessions the teacher instructed the program to select the ten alternative call types. Since these were alternatives, certain features of the original call type were retained, but the progress of the call was often very different. In the following analysis the twenty calls will be referred to by a two-digit number, the first digit will be from the set 1-10 defining the call type number and the second digit from 1-2 defining the original call type (1) which was handled in Periods 1-3 and the alternative call type (2) which was handled during periods 4-6. For the purpose of analysis however, the calls were treated as being independent.

Table 36 shows the average call handling times and error rates for each type of call in periods 1-3 (Session 1) and in periods 4-6 (Session 2). Figures 36 and 37 show these results in graphical form with the calls ordered according to increasing call handling times in the appropriate sessions. Figure 37 indicates the error rates of the calls ordered as in Figure 36 to enable assessment of common trends between times and error rates. Table 37 and 38 are the summary tables for the analysis of variances of call handling times during Session 1 (Table 37) and Session 2 (Table 38). From these tables it is apparent that there are highly significant differences between the times recorded for the different calls in each session ($p \leqslant 0.001$). Tables 39 and 40 are the equivalent analyses of variance summary tables for the error rates committed with different call types in Session 1 (Table 39) and Session 2 Table 40). During Session 1 the error rates for different call types were not significantly different (at $p \leqslant 0.05$) but during Session 2 the error rates were significantly different. This finding is even more important in view of the fact that, as can be seen from Figure 35 overall error rates during the last three periods changed little. This result may indicate that subjects were making very few errors on call types with which they were very familiar by Session 2 but still making large numbers of errors on unfamiliar call types.

Subject groups

The large differences in call handling times and error rates between subjects prompted a deeper analysis to try to determine other measured factors which correlated with performance. Since subjects had been grouped according to eyesight, this factor was first investigated. The average call handling performances of groups A-D are summarised in Table 41 and illustrated in Figures 38 and 39. From analyses of variance, summarised in Tables 42 and 43 it is apparent that the large differences between the performances of the groups were significant (Times at $p \leq 0.001$, Error rates at $p \leq 0.05$). Moreover, from the shapes of the bar charts, it appeared that call handling times and error rates might be correlated.

No sensible causal link could be made between relatively small variations in eyesight and performance on the call-handling task, and therefore another reason was sought for the differences between the subject groups. Since the groups differed significantly in average age this factor is examined in the next section.

Age and call handling performance

In the previous page it was shown that the eyesight groups A-D performed significantly differently during Experiment 5 but no direct link could be ascribed between eyesight and call-handling performance. Since subject groups differed in age as well as eyesight (see Table 32) it was felt necessary to examine a possible link between age and call handling performance. A Kendal concordance test was performed on the ranks of the subject by age and by call handling times, according ranks of 1 to the lowest age and to the shortest call handling time. A Kendall $\tau$ of 0.49 was recorded, a result significant at $p \leq 0.001$, indicating a moderate correlation between age and call handling time. A similar analysis carried out to determine the correlation between age rankings and error rate rankings showed a $\tau$ of 0.37, significant at $p \leq 0.003$, again indicating a moderate correlation between age and error rate.

From the scatter plot of age versus average call handling time (Figure 40) it is apparent that the older subjects generally handled calls more slowly than did younger subjects, but that there was also a wider variation between the speeds of older subjects than there was for younger subjects.

However, since the analysis of call handling performance over successive experimental periods (see the first part of this section) showed that call handling speeds were still improving, it might be taken to indicate that older subjects were learning at a slower rate than were younger subjects but that eventually they could reach equally good levels of performance.

Correlation between performance measures

There were significant correlations between subject age and both average call handling time and average error rate, suggesting a possible correlation between subjects call handling time and error rate. An analysis by ranks showed a positive correlation of 0.5 (Kendal $\tau$, $p \leq 0.001$) between call handling times and error rate indicating that, as error rates increased so did call handling times. This indicates that, subjects who performed quickly tended also to perform accurately, and that their superiority was not achieved at the expense of making more errors.

### D.3.3 Visual Fatigue

The degree of visual fatigue engendered by Experiment 5 was assessed by measuring this before and after the experiment, so that differences could be noted. Two techniques were used to measure fatigue, one measuring subjective visual fatigue with a short questionnaire (Questionnaire 4), the other measuring objective correlates of fatigue using the Mallett fixation disparity test.

Appendix 10 contains the questions and summary of answers to Questionnaire 4. Table 44 shows the numbers of subjects reporting positive or negative changes in visual fatigue symptoms between the beginning and end of Experiment 5. In all, 22 subjects from the total of 28 reported at least one increase of rating in a visual fatigue symptom, although, of these, eight additionally reported at least one decrease in a symptom. Only one subject reported solely decreases in visual fatigue symptoms and five reported no change.

Table 45 shows a further breakdown of the frequency with which changes were reported in each of the six symptoms. Question B4 contained a five point rating scale from 'very watery' to 'very dry' and a change in either direction was taken to indicate fatigue. Each fatigue symptom was reported to have increased by at least four subjects (14%), and thirteen subjects (46%) reported increased tiredness. The symptoms most frequently reported to have decreased were 'blurring vision' (three subjects reported decreases) and "eyelid discomfort" (three subjects). Table 46 shows a breakdown of responses by subject group, indicating that a greater proportion of subjects in Group C reported increases in symptoms than did subjects in Group A, B or D.

Table 47 shows the proportion of subjects in each subject group reporting increases in each fatigue symptom. Ranking the frequencies with which each subject group reported increases in each symptom enabled an analysis to be made on the agreement between subjects as to the most important symptom. A Kendal coefficient of concordance of 0.38 is found which is not significant at $p = 0.05$. Thus, subject groups do not have a significant amount of agreement on the most prevalent symptoms. Moreover, if the subject groups are ranked by increasing proportions reporting each of the symptoms,an analysis shows that differences between the subject groups are not maintained across all symptoms (Friedman Analysis of Variance by ranks not significant at $p \leqslant 0.05$).

In order to assess the general significance of the increased levels of visual fatigue reports following Experiment 5 it was necessary to compare these with equivalent results obtained during a normal working day. All of the experimental subjects who worked at L.M. Ericsson switchboard and who had taken part in Experiment 5 were asked to complete Questionnaire 4 before and after a normal working day. Thirteen sets of reports were obtained and a direct comparison made with their reports before and after Experiment 5, enabling these subjects to act as their own "experimental controls". Table 48 shows the numbers of subjects reporting increases, increases and decreases, no change or decreases in their ratings of visual fatigue symptoms during the day in which they took part in Experiment 5 and during a normal working day. Precisely the same number of subjects reported an increase in their rating of at least one symptom during the experiment and during the normal day (nine subjects in each case).

However, of these nine subjects, four also reported at least one decreased rating of a symptom during the normal day while only two reported both increases and decreases during the experiment. However, as is clear from Table 49, the subjects who reported increased fatigue during the experiment were not always the same subjects who reported increased fatigue during their normal working day. Thus, the increases in fatigue over a day are either due to random factors outside the control of the equipment or are due to extremely complex interactions. Nevertheless, it has been established that, on a population basis visual fatigue is as likely after a day at a conventional switchboard and after a day operating the ANE403 position.

Table 50 shows a comparison between the frequencies with; which subjects reported increases of specific fatigue symptoms during Experiment 5 and during a normal working day. Subjects were at least as likely to report increases in specific eye fatigue symptoms on a working day as during the experiment and, if anything, the results suggest that fatigue symptoms were increased less frequently during Experiment 5 than they were during a normal working day.

Examining the results of the Mallett fixation disparity test applied before and after Experiment 5 (Table 32) indicates that eight of the subjects exhibited fixation disparities after the experiment that were not detected before the experiment. Of the eight, five reported disparities in line with heterophorias detected in the screening test (3 Exophorias, 2 Hyperphorias). However, three reported suppression of one eye, a symptom tested for, but not detected, during the screening test or in the preexperiment Mallett test. The five subjects who were found to have exophoric fixation disparity before Experiment 5 showed no detectable change after the experiment. Although the mere presence or absence of fixation disparity has been shown to be a possible objective measure of fatigue it is believed that in future the extent of the disparity should also be measured. No results have been collected using the Mallett test before and after a normal days work but it was recommended that this should be done if the results obtained during the experiment are to be placed in perspective.

In order to investigate the relationship between visual fatigue and call handling performance,subjects were divided into four groups according to average call handling times (fast and slow) and according to the extent of subjectively experienced visual fatigue. Subjects who handled calls in an average time of less than 150 s were placed in the fast category, those whose average call handling times exceeded this value into the slow category. Subjects who reported increases in at least one visual fatigue symptom and decreases in no symptoms were placed in the high fatigue category. Subjects who reported no changes in any visual fatigue symptoms,or who reported decreases in at least one symptom (even if they reported increases in other symptoms), were placed in the low visual fatigue category. Table 51 shows the frequency distribution of subjects within the four cells of the 2 x 2 matrix formed at the intersection of the two call handling performance categories and the two visual fatigue categories. It is apparent that the faster group more frequently reported 'only increases' in visual fatigue symptoms than the slower group. A Fisher exact probability test established that this frequency distribution could only be expected to occur by chance with a probability of 0.026 so that these results are statistically significant at $p \leq 0.05$.

D.3.4 Assessment of call handling equipment and procedures

Questionnaire 5, shown in full in Appendix 11, was presented to subjects after Experiment 5 and required them to rate the equipment and procedures used in call handling. The frequency with which each answer was given is shown by italic numbers in the appropriate answer boxes, an empty box indicating a nil response.

The answers to Section A, relating to the datascreen itself, suggest that the screen size was acceptable, that the characters were of a suitable size and that they were legible. The brightness and contrast levels were adjustable by the subjects to appropriate levels and very few subjects reported flicker at levels approaching unacceptable. Almost all subjects reported that the image was steady. No-one reported unacceptable reflections of the lights but one subject reported unacceptable reflections of herself, primarily because, as was discovered later, the cleaning staff polished the anti-reflective treatment from the surface of the screen during the period of the experiments. All subjects reported that the colour of the screen was acceptable, with 50% reporting that it was a very good colour.

Section B of the questionnaire related to the format of the ticket display on the data screen, and in this section the wording of the questionnaire was critical if ambiguity was to be avoided. Over 60% of subjects found it easy to learn the positions of the datascreen fields, and a majority (53%) found it easy to learn the orders which had to be entered into the computer, despite a total of only four hours experience with the equipment. Over 60% of subjects reported that it was easy to enter information into the order field but, nevertheless, 75% of subjects reported that they had "often" or "very often" made mistakes when writing orders which they had corrected before "transmitting" them to the computer. Only 50% of subjects had 'seldom' misread or misinterpreted orders or information on the datascreen, but only two subjects out of 28 (7%) reported that this had occurred 'very often'. All subjects found it easy or very easy to enter the A and B numbers into the relevant fields. Only two subjects reported that they found it difficult to find the cursor on the screen, but only one was "often distracted". All but one subject found it easy or very easy to appreciate the state of the subscribers (e.g. engaged), but eight found it difficult or very difficult to appreciate the connection between the operator and the subscriber (e.g. speak to A). This may have been partially caused by programming errors in some call sequences which led to this status being incorrectly displayed.

The responses to Section C, dealing with the keyboard design were generally favourable, although, as in the previous sections, there seemed to be a tendency for subjects to mark the centre point of the scales rather than to consider each question carefully. Most subjects found the physical dimensions of the keyboard acceptable, with only a slight tendency to report that the keys were too small, too closely spaced and too light in operation. There was a wide range of responses to the question about the extent of key travel, with one subject reporting that it was much too little and another that it was much too much. The keyboard slope was found suitable by 86% of the subjects, with two subjects (7%) reporting that it was too much. The layout of the separate sections of the keyboard was 'neutral' or very convenient, but never inconvenient. Relating to the layout of the individual sections, the large majority of users found them 'neutral', convenient or very convenient, but a few subjects reported some inconvenience. The numeric keyboard was criticised in this respect, as were the cursor control keys.

The answers to the Section D relating to the console were also generally positive. A large majority of subjects found all aspects of the chairs very good, although there was some slight criticism of the back support adjustment. The desk was generally acceptable, with scores to rating scales, having two 'unacceptable' ends providing most scores in the centre. Some subjects reported difficulties when adjusting the tilt of the screen, and a number reported that the desk space was insufficient, as was the room for personal belongings. Most subjects found that the general appearance of the desk was 'neutral', acceptable, or very acceptable.

In answers given to Section E, subjects found the lighting a little too bright but generally acceptable in colour. Several subjects considered that the position of the lights was unacceptable but most reported them flicker free and generally acceptable. The acoustic properties of the room were generally judged acceptable, and most subjects were not at all distracted by the sounds of others. The room was judged to be too warm or much too warm by the subjects, but most reported that the humidity was suitable.

PAGE NUMBERING AS ORIGINAL

Sections F and G dealt respectively with the operating procedures of demand and of delay calls. Although most respondents felt that it was easy to appreciate that a call was a demand call, four considered that it was difficult or very difficult. There was a wide range of opinion about the announcing machine, with a slight balance of opinion against it. Twelve people said that they would prefer to make the announcement themselves, and a further nine thought that it would be preferable to be able to select the announcing machines or speak themselves. Only seven people preferred not to make the announcement themselves. This same seven subjects also stated that they preferred to always have automatic acceptance of calls, whereas nine preferred to use the 'accept'-control and a further twelve considered a choice of modes preferable. Almost everyone found it fairly easy to enter the A and B numbers into the appropriate fields,but fewer reported that it was as easy to carry out an A number check. A majority felt that it was very useful to provide country and/or area codes automatically if a name was used and only two considered it of little use. Generally, responses were neutral to the ease of use of a range of special services (Section F3). A few subjects (3-4) consistently reported difficulties with these facilities,but a greater number (12-14) reported that the facilities were 'easy' or 'very easy' to use. Few subjects found it difficult to set up the call to the B subscriber through the various stages, but many responses were fairly neutral. Most subjects were neutral or favourable to the ease with which the operator's position could be 'freed'. The answers to Section G showed a similar pattern, with many subjects choosing the neutral point and few committing themselves to the extremes. The frequency of positive responses was usually 4-6 times the frequency of negative responses.

The answers to questions in Section H show a wide spread of opinions on the relative merits and disadvantages of the new equipment in relation to the operator's present job. Approximately 50% would not commit themselves either way, choosing the neutral response. Generally the response indicated that the majority of subjects found the job more interesting and less repetitive than their present job but somewhat more mentally demanding and tiring. Twelve reported that visual fatigue appeared to be greater and ten reported greater postural fatigue. A large number felt that the job was less prestigious than their present job. Rating the environment, six felt that it was less pleasant than their existing job and eight felt that it was more pleasant.

## D.4 CONCLUSIONS AND RECOMMENDATIONS FROM PHASE III

### D.4.1 Assessment of equipment and procedures

The first purpose of this phase was to check the design of the datascreen, keyboard, console and dialogue in the context of realistic call handling over reasonably appropriate periods of time. Subjects recorded their subjective reactions to the call handling equipment and procedures in Questionnaire 5. The responses to Questionnaire 5 were very encouraging. The datascreen itself was acceptable on all dimensions except that of screen reflections of the operator. This negative finding can be attributed to the fact that the matt finish applied to the screen to reduce its reflective properties was accidentally removed by office cleaning staff during the experiments but not noticed until they had been completed. The ticket layout was generally felt to be good but subjects did report some difficulty remembering order codes, entering them accurately and interpreting computer commands. This problem was due primarily to the short training time available and the limited experience of the subjects in some international types of call used in the experiment. This finding indicates the importance of the training programme. Subjects found it easy to appreciate the state of the subscribers (e.g. engaged) but more difficult to appreciate the state of the connection between the operator and the subscriber (e.g. speak to A). This finding may be related to several programming mistakes, discovered late in the experiments, which caused the symbols which indicated the state of the connection between the operator and the subscriber to be incorrectly set during several types of call. The physical characteristics of the keys and of the keyboard were generally 'acceptable' and the layout of the keyboard was 'convenient'. The seating was generally liked and the adjustments were easy to make although there was a little criticism of the ease of backrest adjustment. The desk was generally acceptable and adjustment reasonably easy. Some subjects felt that desk and personal space was too restricted. The desk space was partially occupied by the paperwork involved in the experiment (questionnaires etc.) and by operator manuals and call handling instruction sheets, all of which would not be required during normal operation. Space for personal belongings must be made available in working operator rooms.

The operating procedures were liked, although there was some criticism of the use of an announcing machine which asked the A subscriber to state his number. Many subjects stated that they would have preferred to make the announcement themselves or at least preferred to be given a choice between the recording and making the announcement themselves. Subjects found it easy to enter the details of subscribers and to check the A-number, but the booking of such special facilities as limited duration calls was considered difficult by some subjects, a finding which can be explained by the limited experience the subjects had of such calls. One of the great advantages of the training system is be the capability to teach operators to deal with infrequent and complex types of call. Once call details had been entered operators found it easy to set up the calls, to cancel them or to book delay calls. Delay calls were very easy to handle as were those to announce 'end periods' or provide 'price advice'.

Rating the job of call handling on the new equipment versus their job operating with conventional equipment elicited some interesting results. Many found it more interesting and less repetitive. It was reported to be equally 'mentally demanding' and 'generally tiring' but several subjects reported somewhat higher visual fatigue (however, this factor is investigated further later in this section). The slightly greater postural fatigue reportedly experienced during the experiment could be attributed to two factors, the tension of being an experimental subject using new equipment and the twisted posture which some subjects adopted when referring to the operator manual or instruction sheets. It was noted that a document holder would be a useful addition to the desk if printed material is provided for operator use. Such a device is considered particularly necessary for trainees.

Observation of subjects before and during the experiment confirmed that people cannot be relied upon to adopt comfortable and healthy postures without help and training. Furthermore, adjustable furniture exacerbates the problem if help is not provided,because subjects chose the wrong adjustment combinations. Most subjects upon being introduced to the equipment did not try many adjustment possibilities and many did not adjust the work-place to any significant extent. When the recommended posture was explained,subjects generally found it easy both to adjust the seat, the desk and the datascreen and to attain a suitable posture. An additional problem arose which had not been envisaged in Questionnaire 5. No mention was made in the literature review

(Phase I) to headset design, but this proved to be a major difficulty with ANE 403 equipment used in these experiments. The problems appeared to arise both in relation to the headset itself and to the plug, which not only provided the audio signal but also switched the operator position "on" when inserted into the socket. Two types of headset were made available; one held on the head by a spring steel band padded at one end and with the earphone at the other to which the microphone was connected by a clip. This design was most satisfactory, particularly to spectacle wearers. The other design, neater and lighter, consisted of a microphone holder which fitted around the back of the ear and to which a tube was fitted on the opposite end of which was a removable earpiece which was supplied in a range of sizes. Some subjects had little difficulty with the device, particularly if they wore no spectacles and if they clipped the lead firmly to their clothing with the clip provided. Spectacle wearers often found the special clip provided to hold the earpiece to spectacle frames difficult to fit and the earpiece difficult to locate. It was recommended that alternative headset designs are provided for operators who should be given ample opportunity to try each time before making the choice. The headset plug is very heavy and has no clip provided enabling it to be clipped to clothing. This encouraged operators to remove the headset when leaving the workstation to take even a short break. Moreover, if the real operator wishes to leave the position during a call, e.g. to refer to a directory, they must choose to remove the headset (potentially very difficult for a spectacle wearer) or unplug the headset from the desk, thus releasing the call. A small in-line socket on the headset wire would solve some problems but must be designed to ensure that the position is closed when the operator is away for several minutes. The main headset plug was large and it was insufficiently clear in which orientation it had to be presented to the socket. This problem was compounded by the fact that the socket was located too far back under the desk and could not be seen by an operator when seated comfortably.

Experimental subjects frequently reported particular difficulties to the experimenter or the teacher and notes were made of these difficulties. Although subjects reported in Questionnaire 5 that the testing of A-numbers was easy they complained about the unnecessarily long and repetitive task. It was considered that it might be feasible to automate this process, completely or partially, and thereby save several seconds per call. One step in this direction which was suggested was to provide an 'OK' button to confirm the positive result of the test, thereby considerably simplifying some keyboard operations. In Questionnaire 5 several subjects reported that they had difficulties remembering and entering order codes with several reporting frequent mistakes. Training was thought likely to reduce these problems considerably but it was also thought that call handling times and error rates might be reduced considerably if common orders could be given from a number of special function push buttons.

### D.4.2 Assessment of call handling performance

The second purpose of Phase III was to provide information on call handling rates, and consequential error rates under realistic conditions. To fulfil this purpose,call handling statistics were printed every twenty minutes over two hours for each subject. These statistics indicated how many calls had been handled, which types of call had been dealt with, the average call handling time per call type and the total number of errors committed on each call type. By necessity, subjects were only given a short training period before the experiment, and consequently most subjects were still learning to handle the more complex calls throughout the experimental period. This meant that the subjects' performance scores were more widely spread by differential learning rates and that the performance scores for more complex types of call were probably more depressed than those of the simpler call types which were learned more quickly. However, the experimenter was able to collect data on the likely final call handling times which operators could be expected to achieve by asking the teacher to handle each of the call types. This also enabled the experimenter to make an assessment of the likely length of operators training by extrapolating the learning curve of call handling times to the expected final performance level.

There were wide differences in the measured performance of subjects but it is not yet known whether these are the results of slower learning or whether they indicate likely differences in final performance. When it was discovered that there were significant differences in the call handling performances of the four subject groups, but no logical connection could be found between eyesight and call-handling performance, a more basic causal effect was sought. It was found that performance was moderately (but significantly) correlated with age, indicating that older subjects tended to handle calls more slowly and with more errors. Although it might be possible that older operators might work more slowly after training it might be hypothesised that a greater effect of age might be to slow learning. Thus, after a longer training period older people might achieve performance figures much closer to those of young subjects. Moreover, it was thought that in real situations the performance of operators might be governed to some extent by the performance of the entire telephone system and that experience in dealing with the wider international network would be of greater importance than it was during the experiment.

D.4.3 Visual fatigue.

It is apparent, from subjective reported to Questionnaire 4 and from objective measurements of fixation disparity, that the majority of subjects taking part in Experiment 5 suffered measurable increases in visual fatigue symptoms over the experimental day. There were some differences between the subject groups, with the myopes reporting greater increased fatigue than the others, but this difference was not maintained across all symptoms. Since the myopic group (Group C) performed significantly faster and more accurately than the other groups, the visual fatigue may have been a measure of generalised fatigue to higher workload rather than a peculiarly ocular phenomena. To assess the general significance of the increased visual fatigue, thirteen subjects completed Questionnaire 4 before and after a day working on a conventional switchboard and these results were compared with their results during Experiment 5. A very similar pattern of responses was achieved in the frequencies with which subjects reported increased, decreased or no changes in visual fatigue symptoms; suggesting that, over the group, operating ANE 403 equipment is no more visually fatiguing than operating a conventional switchboard. Examining individuals, however, indicated that some experienced greater fatigue operating the ANE 403 position and others experienced greater fatigue with the conventional switchboard. No underlying causal factors have yet been found for this effect, and it cannot be predicted by the age, eyesight or call handling performance of subjects.

# CHAPTER E - DISCUSSION AND CONCLUSIONS OF THE STUDY

## E.1. PREFACE

The previous three chapters have described the major phases of the project, indicating how ergonomics knowledge and methods were used to influence a major product design. Chapter B described Phase I, in which past knowledge contained in published papers and reports was used to identify problems, to provide answers and to guide further work. Chapter C reported on a series of short, interlinked, problem-oriented experiments used principally to identify the most effective datascreen design and to examine ambient light levels further. Once prototype workstations were available and linked to a training system it was possible to conduct Phase III (see Chapter D), in which the entire operator interface was tested under reasonably realistic handling conditions. Now that the first ANE 403 system is in operation, and it is understood that several more have been ordered, it is hoped that the final phase of the recommended plan for the project will be conducted (see A.8). Phase IV will consist of ergonomics tests and measurements upon a working ANE 403 system after it has been in service for several months or years, and is properly outside the timescale of this thesis. The purpose of the phase will be to identify remaining problems and to assess the success of the product. This success will be measures in terms of the performance of the system and its acceptability to both the operators and the administration. Consideration will be given to solving the problems on the installation studied and a report will be produced to ensure that future designs benefit from the field study.

The purposes of this, the final chapter in this thesis, are to describe briefly how the ergonomics work reported in previous chapters influenced the final product design, to describe the direct and indirect benefits which accrued from the ergonomics work and finally to identify the salient features of this product which identify it as a model of its kind.

## E.2. ERGONOMICS ASPECTS OF THE FINAL DESIGN

The majority of the ergonomics recommendations made in the literature review report (see Phase I) were adopted with certain modifications made as a result of Phases II and III. The hardware of the datascreen, the keyboard and the console followed the ergonomics specifications closely, except where minor changes were made after consultation with the ergonomist. The layout of the datascreen format and the dialogue involved in the call handling procedures followed the ergonomics advice closely and a few changes were instigated as a result of Phase III. The recommendations relating to the lighting and other aspects of the operator room were passed on to purchasing administrations, together with the evidence upon which the recommendations were based.

### E.2.1. Datascreen

The datascreen was chosen to conform to the ergonomics specification provided as a result of Phases I and II (see Section B.2 and Chapter C). Figure 41 shows a view of the terminal illustrating the outer case, hood, screen and keyboard. Figure 42 is a closer photograph of the datascreen face showing the displayed ticket designed by the author (see Fig. 9). The phosphor chosen was P39 following the original recommendation made in the report of Phase I because the most effective datascreen and treatment tested in the experiments of Phase II, P31 phosphor plus a matt polaroid filter, could not be manufactured in the recommended form (i.e. with the filter moulded to the face of the screen). The datascreen was not fitted with any form of filter but was treated with an anti-reflective coating to reduce the mirror reflections of the operator. The shape and size of the datascreen characters (seen in Fig.42 ) followed the recommendations precisely (see B.2.1). For example, it was found that the suboptimal character width/height ratio of 0.66, originally considered necessary for space reasons, could be modified to produce the ideal of a 0.75 ratio by adjusting the overall height of the datascreen format and effectively compressing the picture so that it no longer filled the datascreen. The idea of using an electronic means of character rounding (see B.2.2) could not be adopted, because the existing integrated circuit 'chips' were made for viewdata and teletext applications which used a 5x7 dot matrix not a 7x9 matrix as used for ANE 403. The character shapes recommended by the author (see B.2.3) were programmed into the programmable read only memory (PROM) ensuring an easily distinguished character set. In Phase II it was established that the relative brightness of the normal and high brightness characters were too widely different, making it difficult to set up an appropriate brightness and contrast setting suiting both brightnesses. This problem was eliminated in the prototype used for Phase III and in the final product.

E.2.2 Keyboard

The keyboard designed specifically for ANE 403 followed the ergonomics recommendations exactly and is shown in Figure 43. The keys used were those recommended by the author and they proved acceptable with small reservations about their 'feel' and their rather noisy operation. The colour coding of the keys, suggested in Section B.3.3 (indicated in Figure 11), was disliked by some company engineers, who thought colour unnecessary and somewhat ugly, preferring uniformly black keys. However, it is understood that at least one potential customer would prefer the use of colour coding on the keyboard and this option will be provided. The slope of the keyboard, recommended to be 20° in the literature review (Section B.3.5), was finally decreased to 10° at the suggestion of company physiotherapists, who reported that they had found wrist strain in operators using steeper keyboards (see Figure 41 ).

E.2.3 Console

The ergonomics recommendations for the console, contained in Section B.5 of the Phase I report, met with some opposition from two different and mutually opposed groups within LME. One group of people thought that adjustment was unnecessary and would be misused, the other considered that every major surface should be independently adjustable to ensure optimal fit. At the recommendation of the author, the company physiotherapists carried out an extensive test of three alternative console designs in order to resolve this conflict:-

- Fully adjustable. Independent adjustment of the heights of the chair, datascreen and keyboard.

- Limited adjustment as recommended by the author. Adjustable chair, and joint adjustment of datascreen and keyboard.
- Non-adjustable console with adjustable chair.

Fifty subjects of widely varying stature were used in fitting trials to evaluate these three designs (see Figure 44). Two experimental conditions were used:-

- Subjects adjusted the chair and console (if adjustable) without assistance.
- Subjects adjusted the chair and console with help from a physiotherapist with appropriate ergonomics knowledge.

Briefly the results can be summarized as follows:-

- The non-adjustable console could not accommodate the tallest or the shortest operators without exceeding the ergonomic criteria (i.e. viewing distance, elbow position, etc.).
- The subjects were seldom capable of adjusting the consoles appropriately without help and also it was clear that the greater the number of adjustments allowed, the more lengthy would be the training necessary to use them properly.
- The console with a limited height adjustment (i.e. the datascreen and keyboard moved together) enabled the console to be fitted adequately to all subjects without exceeding the criteria.

As a result of the study outlined above it was recommended that, despite the more exact fit possible with a fully adjustable console, the difficulties of making the necessary adjustment militated against such a complex arrangement. It was thus decided that the partially adjustable console originally recommended in Phase I would be used.

Figure 45 shows the final configuration of the console. The desk surface, supporting both the datascreen and the keyboard was capable of easy adjustment between 600 and 700 mm. Within this range the desk surface was sprung upward against the weight of the equipment and was adjusted up or down by a hand-crank driving the desk, via a worm gear-pinion link, against friction. The gear drive was chosen to be unidirectional in operation, effectively preventing the mechanism from being driven in reverse should an extra load be applied to the desk. A linear rule was incorporated into one side of the console, providing a quick check on the height of the workstation and enabling each operator to quickly select their normal height (see right hand edge of desk in Fig.41 ). In addition to the range of easy adjustment described above, the screw-jacks incorporated in each leg of the console enabled the desk surface to be adjusted to any height between 560 and 760 mm as recommended.

E.2.4 Dialogues

The dialogue suggested in Phase I was found to be easy to use in the experiments of Phase III and was therefore adopted in the final design. Certain procedures were found somewhat tedious (e.g. acknowledging a request from the computer by typing OK in the order field) and it has been suggested that in future these may be simplified by the use of special function buttons.

An example of a call-handling dialogue based on the author's work is shown in Appendix 12. This is taken from part of the operator training manual produced by LME to accompany ANE 403.

E.2.5 Lighting

In Phase I the importance of well designed room lighting was stressed, and recommendations were made for the control of light level, direction and flicker. It was also pointed out in Phase I, that regular maintenance and lamp replacement was essential to ensure that lighting conditions did not deteriorate excessively from their initial values. In Phase II it was found that, by careful design of the datascreen and console, and by the selection of appropriate lighting louvres, the illumination level could be increased to 500 lux without deterioration of the operator's comfort or performance.

These recommendations have been passed to potential purchasers of ANE 403 so that they can taken them into account when meeting the architectural requirements for the system.

E.2.6 Work levels and rest periods

In Phase I, it was noted that the task of the ANE 403 operator was likely to be visually demanding and would involve restrictions upon physical movement. Thus, it was thought that regular and relatively frequent rest pauses would be necessary, to enable visual and postural fatigue to be dissipated. In Phase II visual fatigue was measured and found to increase significantly (in statistical terms) between the beginning and end of the experimental periods. However, this strain was reported under deliberately poor lighting conditions, designed to test the effectiveness of screen treatments. Moreover, since no control group was used in Phase II, it was not possible to assess the differences of fatigue between working with ANE 403 and performing another task for the same periods. Phase

III was used to evaluate the entire operating system under more realistic conditions of lighting and work pressure. Since statistically significant increases of visual fatigue had been reported in Phase II, opportunity was taken to measure visual fatigue and postural fatigue subjectively and to attempt to measure visual fatigue objectively. The results can be summarised by stating that reported increases of visual and postural fatigue over the experimental periods were generally not significantly different from those reported over the normal working days of operators of a conventional exchange.

It is hoped that in future more extensive study can be made of methods for assessing visual fatigue and that these can be used in extensive experiments to determine the most effective work-rest periods. Indeed, directly as a result of the findings of this project research funds have been obtained from the European Coal and Steel Commission of the European Economic Community to investigate the extent and possible causes of visual and postural fatigue at visual display units.

E.2.7 Work organisation

Considerable emphasis was made in Phase I of the importance of work organisation as a factor which would influence operator performance directly and indirectly (via dissatisfaction, absenteeism, staff turnover, etc.). However, since work organisation is determined by the operating administration and no field sites were available for study no further work was done in this field. However, the Phase I report is being passed in its entirety to purchasing administrations who are therefore being made aware of the importance of this issue.

The importance of providing interesting room layouts was raised in the Phase I report, and, as a result, the LME Company architects prepared a series of alternative desk layouts for hypothetical rooms for the benefits of purchasing administrations. Two examples of possible room layouts are shown in Figures 46 and 47.

## E.3 DIRECT BENEFITS OF ERGONOMICS INVOLVEMENT

This section will consider the benefits arising from the ergonomics study which directly influenced the effectiveness of the ANE 403 system for three groups of people:-

- Purchasing administrations
- Operators
- Subscribers

The effectiveness of the ANE 403 system will ultimately be determined by its impact upon all of these groups, and thus the author was forced to treat each of them as customers for his services.

### E.3.1 Benefits for purchasing administrations

For a manufacturer such as LME the purchasing administration (such as the Post Office in Great Britain) represents the principal customer who must be convinced if a sale is to be made. However, a complex piece of equipment such as a telephone exchange must be sold on its potential performance not simply its technical quality. Telephone administrations have always realised in a general way that system performance depends heavily on the human element. They are now beginning to realise the crucial role that ergonomics can play in solving specific problems which, if not overcome, may render telephone networks and exchanges ineffective, and at worst inoperable. It is now recognised that, not only will individuals perform badly if the man-machine interface is poorly designed or if their work is poorly organised, but that the workforce may collectively influence purchasing decisions and operating procedures. Moreover, consumer bodies are becoming increasingly powerful in representing the needs of subscribers. Thus, there is a growing recognition by equipment manufacturers that they must sell their products not only to the administrations directly, but also to the final users, be they operators or subscribers (see Sections E.3.2 and E.3.3). There is increasing evidence that all of these users are becoming aware of the need for effective use of ergonomics in datascreen design. As witness to this fact many of the major Trade Unions in Britain and throughout Europe have drawn up guidelines for the purchase, installation and use of Vdu's in offices and elsewhere (e.g. ASTMS, 1979).

The ergonomics advice provided by the author should ensure faster and more accurate call handling of operator assisted calls than would have occurred without its benefits. These improvements are expected to arise from the application of basic ergonomics knowledge to information formatting, keyboard layout and operating dialogue to ensure that they match the information processing needs and capabilities of the operator. However, it is important to note that a reduction of call-handling time is only a single criteria by which performance may be judged and certain recommendations contained in the study may increase call handling-time in the interests of improving quality in other ways. Two examples of such an approach can be taken from the field of work organisation.

It was suggested that operators work in functional groups rather than as individuals, slightly reducing the flexibility of the system in the interests of operator satisfaction. Another suggestion was that administrations should be able to provide a facility to operators whereby delay calls are permanently assigned to the operator who handled the original demand call. Again this facility reduces system flexibility and may delay certain calls but it is hypothesised that under some circumstances it could lead to increased satisfaction for operators and subscribers.

Operator supervision is an important tool of control for administrations,but the way in which such control is exercised varies widely in different cultures. The new electronic exchanges can provide faster and more comprehensive statistics on individuals and groups of operators,as well as for the entire exchange. Such information is a vital element in the automatic scheduling and call distribution system, enabling it to assign incoming and stored calls to operators in such a way that work load is distributed easily. However, there is a danger that easily measured variables (e.g. call handling times) will be used unjustly because other equally important variables (e.g. error rate) cannot be measured easily and there may be no breakdown call types. In addition, there are severe dangers which can arise when supervisors come to rely upon electronic surveillance because personal face-to-face contact with operators is excluded by the way work is organised and individuals are seated.

Other direct benefits to the administration of the involvement of ergonomics can be found in areas which are frequently overlooked. Operators are usually treated as semiskilled workers and a fairly high degree of lateness, absenteeism, illness and staff turnover is expected even if it is not openly tolerated. It is important to note that many of these indices are strongly influenced not only by comfort but by job satisfaction and motivation which in their turn are determined by work organisation (e.g. autonomy, social contact, performance feedback). Thus, in the design of a complex system, it was essential that the ergonomist considered ways in which he could influence this work organisation, directly by designing tasks which require human skill and experience and indirectly by informing purchasing administrations of some of the pitfalls and possible solutions.

When designing the equipment, tasks and work organisation the author tried to improve the overall quality of the operator's working life. Ultimately these benefits will be transmitted to the administration by a reduction in the hidden costs of staff training, overtime and social costs.

E.3.2 Benefits to operators

Since operators directly use the system and since the majority of the project was devoted to the design of the interface between the operator and the computer system one would expect the ergonomics advice to benefit the operator.

The first direct benefit should come in terms of the increased operator comfort and reduced visual and postural fatigue resulting from a workstation design which complies with ergonomics recommendations. In the long term these benefits may reduce some of the health problems associated with inadequate workplaces (e.g. back pain). The adjustment of the seat and of the desk surface will ensure that a very wide range of people can be accommodated comfortably, with the datascreen and keyboard in the correct position for the operator's eyes and hands respectively.

The use of a datascreen terminal linked to the computer enables the operator to enter information onto an electronic 'ticket' once-and-for-all, thus eliminating many of the repetitive re-dialling required in previous exchanges. Whenever a call is being handled the information relating to it is clearly and consistently displayed and re-dialling requires the operation of a single key. Delay calls booked for a certain time are automatically administered by the computer and presented to a free operator when the appropriate time is reached.

However, the benefits to operators described above are primarily 'hygienic', that is, they are concerned with the minimisation of fatigue and the maximisation of performance using traditional ergonomics techniques. However, there are other problems connected with job satisfaction, which do not arise solely from attention to such 'hygiene' factors. These problems are primarily psychological in nature but their effects may easily swamp the benefits arising from attention to the technical aspects of the operator's job.

In the earliest telephone exchanges, operators were locally based and knew many of the subscribers. Thus, operators played an important social role and thus acquired some recognised status. With increasing automation of the telephone system and a decreasing reliance upon the operator as a facilitator of communication, the social role of operators has diminished to almost zero. As operators have been centralised and few subscribers ever know where they are located there has been an increased emphasis on technical efficiency (speed, low error rate, etc.) and operators are now required principally to translate a spoken request by the subscriber into a computer input. As technology is used to replace experience and skill, and thereby to make operators predictable and alike, it is not surprising that motivation and sense of responsibility decrease with a consequent danger of performance decrement. Thus, ergonomically better and technically more advanced call-handling systems do not necessarily, or of themselves, result in better call handling performance or increased operator satisfaction. Several authors have demonstrated, for example, that the operators of the older cord switchboards may perform better than those of the newer 'cordless' switchboards (Harris, 1975) and that operators of the cordless boards are less satisfied with their jobs (Habberfield, 1971). Brown _et al_ (1980) point out that despite the human factors applied to the cordless switchboards during the 1950's and '60's their job had lost skill variety, task identity, task significance, autonomy, performance feedback and social interaction. Indeed, Brown _et al_ (op.cit) point out that the increased comfort, spaciousness and quietness of the newer switchrooms have exacerbated the operator's feelings of remoteness. Brown _et al_ (op.cit) found that hardware and procedural constraints of the system inhibit operator skills in providing customers with a high-quality service. Brown _et al_ suggest the need to reverse certain task and equipment design trends to restore operator's job satisfaction to previously high levels.

Aware that many of these problems might arise, perhaps to a greater extent in the new operator positions, the author attempted to influence the design in three ways. First, he tried to build into the dialogue a degree of operator flexibility. For example, the ticket completion format allows operators to enter the information in whichever way seems appropriate. Secondly, he persuaded the manufacturers to offer the facility to the administrations whereby operators could choose to hold calls to their position for supervision or delay call handling. Thirdly, he impressed upon purchasing administrations the importance of organising and supervising the working groups in such a way that social contact could be formed and maintained among operators and between them and the supervisor.

E.3.3 <u>Benefits to subscribers</u>

Ultimately the entire exchange system is there to serve the needs of subscribers. The electronic storage and automatic processing of the ticket information, and the consequent capability to use any 'free' operator to handle a call, should provide subscribers with faster, less error prone and ultimately cheaper operator services. For example, calls made directly by the subscriber should be answered more promptly than at present, because each call joins a multiserver queue. Moreover, booked calls will be fed to an operator automatically, ensuring a more precise call time than would usually be possible on a manual system. Finally, the automatic dialling facility available to the operator will enable calls to be handled faster and with fewer errors.

Apart from the direct benefits to subscribers outlined above, it is hoped that the benefits to the comfort and well being of operators will also improve the service to subscribers. Less fatigued and less stressed operators will hopefully be more pleasant and effective.

## E.4 INDIRECT BENEFITS OF ERGONOMICS INVOLVEMENT

In addition to direct benefits expected from the ergonomics involvement, in terms of the quality and performance of the product, other indirect benefits have already accrued.

### E.4.1 Effective design decisions taken

It is often claimed that ergonomics information is obtained only at considerable additional expense even when the cost of implementing the advice is excluded. However, the author claims that it costs little, if any, extra to make good decisions rather than bad ones. Many of the firm recommendations made by the author were adopted without change, saving the company the cost they would have incurred in making the decisions. Moreover, it is held that had faulty decisions been made based upon inadequate ergonomics information the costs of remedying or living with them might have been considerable.

### E.4.2 Change in attitude

The second indirect benefit of the ergonomics study was the considerable change in attitude which took place within the engineering project team towards the role and usefulness of ergonomics and a growing interest on their part to considering the final user in design. This change in attitude spread beyong the immediate project team and several other departments of LME obtained ergonomics advice for the products they were designing.

### E.4.3 Sales benefits

Designing an effective product is only the first step towards successful commercial enterprise. The product must also be marketed effectively, especially in a highly competitive industry such as telecommunications. Although the ergonomics advice was not initially sought to promote sales directly, changes in the public awareness of ergonomics and particularly concern about VDUs made the ergonomics work a crucial aspect of the sales campaign (see Appendix 13). As the ergonomics aspects of VDUs became more important during the development of ANE 403, LME gained a psychological advantage as well as a time advantage over competitors, because the ergonomics advice was built into the development programme from the beginning. It is noteworthy in this context that no evidence was presented during the years of ANE 403 product development which fundamentally altered any of the logic on which the product design was based. Thus LME were effectively placed at least 2 - 3 years ahead of many of their competitors in terms of ergonomics influence on the design of their product. It may be noted, however, that the continuous association of the author with the company during the development years would have enabled new ideas or evidence to be made available and design changes to be considered.

### E.4.4 Effects on other product designs

In the years during which ANE 403 was developed other VDU based product programmes began at LME. The teams responsible for these programmes were able to adapt large parts of the ANE 403 recommendations for the design of the products for which they were responsible. It is a measure of the generalisability of the recommendations that they required few changes to be used for a variety of other products.

## E.5 SALIENT FEATURES OF A MODEL PROJECT

Several features of this project mark it as a model of its kind. These features will be described in this section.

### E.5.1 Early consultation

The first important feature is the early consultation with an ergonomist. The company identified the need for ergonomics advice at the earliest conceptual stage of the project and promptly requested a literature review of the relevant ergonomics factors. The literature review not only enabled evidence to be obtained on which to base design decisions but also enabled plans to be drawn up for the further involvement of ergonomics advisors to tackle unresolved problems.

### E.5.2 Detailed project brief

The project brief, prepared in conjunction with the author and his supervisor, specified in detail the areas which would be covered, and stressed the need for a solution oriented approach rather than an analytic approach. The engineers at L.M. Ericsson had produced preliminary designs for the datascreen format, the keyboard, and the dialogue, enabling the consultants to become quickly aware of the requirements for the future system and the capabilities of the supporting computer equipment.

### E.5.3 Limited problem oriented experiments

The literature review provided useful information relating to the design and use of datascreen terminals and associated equipment, based principally upon published work. However, some design decisions could not be based upon this literature alone. Sometimes this was because no appropriate quantitative data or theoretical model existed to create the data. However, in most cases the difficulties arose as a result of conflicts between several principals. For instance,increasing the brightness of a datascreen character could increase contrast but at some indeterminate stage might cause blurring. It was impossible to tell how any particular equipment would behave without carrying out tests of some kind.

A limited series of brief experiments were conducted to decide on the most effective design decisions, relating particularly to the design and treatment of the datascreens and the appropriate levels of ambient illumination under which it would perform adequately. These experiments were designed to answer specific design decisions and not analytically to explore the behaviour of all possible treatments.

### E.5.4 Adoption of recommendations

The recommendations made by the author were often closely related to each other,forming a network of self-supporting design decisions. Thus, if these recommendations had been adopted piecemeal,their individual effectiveness could have been lessened, thereby considerably reducing the overall ability of the system to meet ergonomics requirements. However, the recommendations were wholly adopted and the final design follows them precisely.

### E.5.5 Prototype testing under realistic conditions

The total development period of the project was of the order of four years, and it was essential that at significant stages there were checks upon the design, since the cost of correcting errors tends to increase at successive development stages. It was important to realise that in the early stages of the project most design decisions were based on limited evidence and their relationships could only be hypothesised. It was therefore considered important to check the effectiveness of the entire system under reasonably realistic call handling conditions.

## E.6 POSTSCRIPT

It may be noted that the original reason for adopting a screen based operator position was to enable call details to be stored in a computer compatible form for later processing by the accounts department. The screen based system provided several benefits for the operator, but it was recognised that these benefits would only accrue if the ergonomics aspects of the system were well considered. In the event, the author was able to provide a detailed design brief; covering not only the terminal design but the physical and social environment within which it was to be used.

This reliance upon an ergonomist to produce detailed design guidelines places a great deal of responsibility upon the ergonomist if the system is subsequently found to fail. The terminal and console have been a great success, and have subsequently been used in several other LME products with only minor changes.