

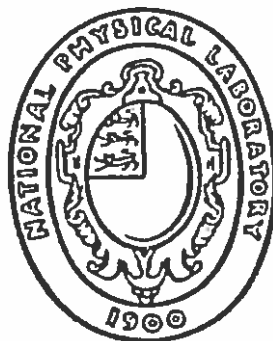
NATIONAL PHYSICAL LABORATORY

AUTONOMICS DIVISION

AN INVESTIGATION OF AUDIO OUTPUTS
FOR A READING MACHINE

by

P. W. Nye



A Station of the
Department of Scientific and Industrial Research

Crown Copyright Reserved

Extracts from this report may be reproduced
provided the source is acknowledged.

The work described above was carried out in
the Autonomics Division of the National Physical Laboratory
on behalf of St. Dunstan's. This paper is published by
permission of the Acting Director of the Laboratory and
with the approval of St. Dunstan's.

PLEASE NOTE

A gramophone record is enclosed, in an envelope on the
back cover, containing examples of the audio outputs
described in the text.

Approved on behalf of Director, NPL by
Dr. A.M. Uttley, Superintendent of Autonomics Division

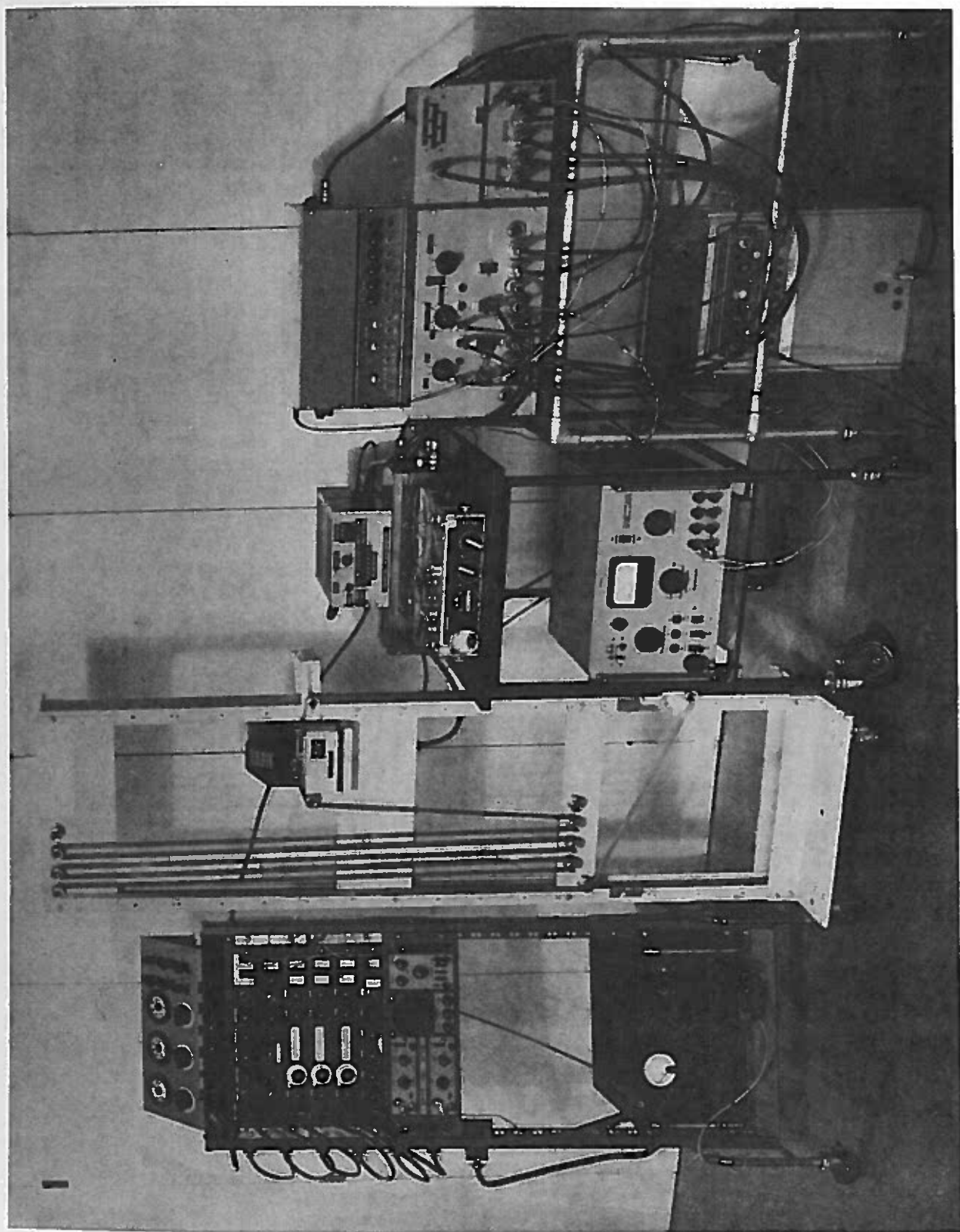
AN INVESTIGATION OF AUDIO OUTPUTS FOR A READING MACHINE

by

P.W. Nye

ABSTRACT

It is intended that this report should draw together the most important results of the work carried out during the three year period from November 1961. The particular topic of this study forms only one of a number of problems requiring attention before a practical reading device can be built. At the time when this research began the general problem of coding auditory information to achieve higher reading speeds appeared to be the most important. Some definite progress has been made in this field and the prospects of further improvements are good, but the investigators have, during the course of the research, become aware of a number of factors which might make future progress more difficult. The report discusses a number of these questions. In this context the writer particularly wishes to acknowledge many helpful discussions with colleagues and members of the St. Dunstan's Scientific Committee.



APPARATUS FOR GENERATING THE EXPERIMENTAL SOUNDS FROM SIGNALS ON PAPER TAPE.

AN INVESTIGATION OF AUDIO OUTPUTS FOR A READING MACHINE

Chapter 1

- 1.0 Introduction
- 1.1 The reading requirements of Blind People
- 1.2 The range of potential reading machines
- 1.3 The best solution for the future
- 1.4 Topics studied in this Report
- 1.5 Possible effects of an uninflected output

Chapter 2

- 2.0 The extent of theories of human communication
- 2.1 The span of immediate memory and reading speed
- 2.2 Response time of the human ear
- 2.3 Redundancy in print and its role in minimising error
- 2.4 Redundancy and dimensionality in speech
- 2.5 Acoustic dimensionality permits higher information transfer
- 2.6 Possible articulatory influences in speech perception
- 2.7 Reasons for examining the feedback theory of speech perception
- 2.8 Methods of testing the feedback theory of speech perception
- 2.9 An alternative hypothesis

Chapter 3

- 3.0 Introduction
- 3.1 Considerations governing the choice of dimensions
- 3.2 The five dimensional and three dimensional codes
- 3.3 Conduct of the experiment
- 3.4 An analysis of the results and conclusions

Chapter 4

- 4.0 Introduction
- 4.1 A mechanical scanning system
- 4.2 The feature analysis program
- 4.3 The token matrices
- 4.4 Calculation of a data reduction factor
- 4.5 Assembly of the stimulus generator
- 4.6 The simulated battelle optophone
- 4.7 The multi-dimensional optophone
- 4.8 The parametric artificial talking machine

Chapter 5

- 5.0 The Haskins training method
- 5.1 The NPL training method

Chapter 6

- 6.0 The conventional and compressed optophone outputs
- 6.1 Conduct of the experiment
- 6.2 Results and conclusions
- 6.3 The similar and dissimilar PAT outputs
- 6.4 Conduct of the experiment
- 6.5 Tests under noisy conditions
- 6.6 Results and conclusions
- 6.7 Generation of 8 word vocabulary PAT MDO and Wuhzi outputs
- 6.8 Conduct of experiment
- 6.9 Results conclusions and discussion
- 6.10 Generation of synthesised Wuhzi and VVO vocabularies
- 6.11 Conduct of experiment using these outputs
- 6.12 Conclusions
- 6.13 Effects of training procedure on learning
- 6.14 Conduct of experiment with PAT, MDO and VVO outputs
- 6.15 Generation of stimuli to test discrimination hypothesis
- 6.16 Conduct of Experiment
- 6.17 Discussion of the results

Chapter 7

- 7.0 The role of introspective observations
- 7.1 Minimising learning effort and
- 7.2 Some possible difficulties
- 7.3 The principle factors controlling speed of audio communication
- 7.4 Future developments

AN INVESTIGATION OF AUDIO OUTPUTS FOR A READING MACHINE

Chapter 1

1.0 Introduction

The need for research into reading devices for the blind arises because the cost of producing braille books is high and consequently the number of titles available is limited. Furthermore, braille books are bulky in comparison with printed books and take many highly skilled man-hours to encode and produce. The idea of avoiding this dilemma by producing a reading machine which would convert ordinary print into an auditory or kinesthetic-tactile display has been under study at various times during the past 50 years. It is often assumed that such a device, if made cheaply enough for personal ownership, would free many blind people from embarrassment by permitting them to read personal typewritten correspondence, and would also enable them to read books and daily newspapers. In particular, the needs of the blind student wishing to read specialised books not available in braille are also raised as justifying this work.

1.1 It must be admitted, however that at the present time there is very little factual information about the reading needs and abilities of blind people which could justify these assumptions. The view is held by many people that the desire among the blind to read newsprint and correspondence is less than is popularly believed, and that making access to book print easier should be the prime objective. Furthermore, if this is true, the proper approach would be to improve the availability of braille by working to reduce bulk and production costs and increase the speed of translation. This argument is partly supported by current opinion of the potentialities of reading machines. It is clear that the types of reading machine that could be produced sufficiently cheaply for personal ownership are likely to permit, even with extended training, reading speeds significantly below the levels achieved with braille, particularly by those trained to read braille from early childhood. This essentially is the case against work on reading machines. In its favour is the argument that to give blind readers access to the same reading matter as sighted persons is easily the most elegant solution both administratively and socially, despite its other limitations. Moreover, in the absence of clear evidence on the matter, it is quite reasonable to make assumptions about the reading patterns of the blind on the basis of the experiences of sighted readers. These would suggest that even a limited performance device could meet the needs of a sizeable proportion of the blind population; probably about 4% (16). It is hoped that in the near future the Ministry of Health will accept a proposal which has been prepared by representatives of St. Dunstan's and the M.R.C. to sponsor a survey of the current reading habits of blind persons in England and Wales. Meanwhile, in the absence of this human data on which to base research objectives, economic considerations usually predominate.

1.2 There are broadly two kinds of potential reading machine. The first is the direct coding type which converts printed text, scanned by a photoelectric sensing device, into sound patterns bearing some relationship to the letter shapes. The reader has to learn to recognise the sounds and to associate them with the printed letters. The second possible machine differs from the first in that the machine itself is responsible for recognition at the letter level and could be made to produce spelled speech sounds or connected speech at the output. Both the recognition machine and the direct coding reader would be required to function satisfactorily with a wide range of print style and quality variants. The technical problems of designing the recognition machine for this standard of performance will take many years to solve and in the initial stages the machine will certainly be costly and cumbersome. But the machine would have the advantage that its output could be made to closely resemble ordinary speech and avoid many of the difficulties of learning and the slow reading speeds that beset the simpler reading devices.

1.3 It is quite clear that a machine producing ordinary speech will be the best solution to this problem, but at the moment the economic and technical difficulties are such as to make it unlikely that a machine can be produced for personal ownership for many years. It is possible that, as an interim measure, some modifications could be made to the direct-coding type of machine which would enable it to partially meet the demand until a more satisfactory machine is made available.

1.4 A considerable amount of work has been carried out in the U.S.A. and U.S.S.R. on variants of the Fournier d'Albe optophone, which is the classic direct coding machine - see Appendix A. The results of all these recent investigations have strongly confirmed that the system has inherent limitations which restrict the reading speed to between 10 and 20 words per minute (wpm) for even the best pupils. Cooper(5) and Beurle(4) have suggested reasons for the poor performance of the optophone and the observations discussed in Section 2.0 incorporate a number of their conclusions. The present study has been concerned principally with the audible display and has concentrated on examining the effects of modifying three characteristics of the optophone output. These have been suggested, by earlier studies, to be the underlying cause of its poor performance.

Firstly, the optophone transforms two dimensional characters into the acoustic dimensions of frequency and time. Experiments have been carried out to examine the effects of increasing the number of dimensions.

Secondly, printed characters contain a considerable amount of visual redundancy which appears in the optophone output in a form which cannot be utilised by the ear. The effects of removing some of this redundancy from the display has been explored.

Thirdly, a widely held theory of speech perception holds that better discriminations can be made among certain speech sounds than non speech sounds. This suggests that to achieve the best results from the human ear the output of a reading machine should be mimicable.

An apparatus incorporating all these modifications has been simulated and the results of subjective tests compared with those of earlier workers. This device, which is first and foremost a research tool, lies somewhere midway between the direct coding and the recognition machines. To create a practical machine considerable reorganization of the design and some further work on the basic concepts will be necessary, but the results of this research show that progress in the direction of a compromise between direct coding and recognition can be made. The outstanding question will be to what extent design simplicity must be sacrificed to achieve adequate reading speeds.

1.5 The three factors considered in this study are all features which distinguish the outputs of direct coding devices from speech. There is a fourth factor which can certainly affect performance. In speech, the structure of a sentence can be conveyed by variations in the pitch and stress of the voice, whereas the reading machine effectively talks in a monotone. In the case of sentences having a simple structure this is unlikely to be a serious disadvantage but in the case of more complex constructions greater difficulties are likely to arise. Sentences of the type, "The family, the woman we met yesterday told us about, is leaving tomorrow.", have, in Yngve's(21) terms, a 'regressive structure' in which the subject of each clause has to be memorised until its predicate is reached. The stress and pitch of the speaker's voice customarily gives the listener information about the location of the clauses which would be almost entirely absent in the output of a reading machine. The effect of this loss of information can only be a reduction in reading speed. Insufficient time has been available in this study to find out the likely loss of efficiency with different kinds of reading matter but it seems probable that book print would suffer most severely.

Chapter 2

2.0 Background Review

The problem of specifying a good design for a reading machine can be neatly stated as that of finding an optimum transformation of printed information from the visual to the auditory modality. Expressed in this way, the problem appears precise and deceptively 'straightforward', but it raises a large number of unanswered questions. A man reading aloud is carrying out such a transformation, but natural speech may not be the only auditory display that could facilitate this optimum performance. There may be many other ways in which information from the printed page could be optimally transformed. To be certain that other methods exist, we would require a general working model of the human processes involved in the evolution of alphabets, in reading and in speech. With this knowledge, we could hope to deduce whether or not the class of displays called speech are alone the most efficient means of achieving the transformation and, if so, precisely how much poorer a performance one should expect from non-speech displays. However, little or none of this relevant data exists and, in the absence of such information, only limited attempts toward developing some kind of model have been possible. Theories in this field and some of the 'facts' are in many respects speculative and incomplete, and because a rapid solution is required as an interim measure the approach to the work has been only in part pragmatic and in the remainder intuitive. However, where some concept has influenced certain activities and objectives in this study, an attempt has been made in the following sections to describe the relevant parts of this model.

2.1 One reason why reading with an optophone is so slow lies in the response time of the human ear. When the repetition rate is low, clicks and other short stimuli are heard as separate events; as the rate increases, the sound changes to a buzz at about 20 cps and then to a tone of rising pitch. Even if the stimuli are not the same, their individuality is lost at 20 cps. This indicates that there is a maximum repetition rate for the reception of discrete stimuli which is limited to about 15 per second. The figure can explain the limiting rate for the reception of Morse code. On average, Morse symbols contain approximately three elements per letter and, taking the average word length in English to be $4\frac{1}{2}$ characters, gives an estimated maximum of 65 words per minute, which agrees very closely with the best recorded reading speeds with the Morse code. A similar calculation for the optophone based upon the need to discriminate five chords per letter gives a somewhat lower figure, again agreeing with the best reading speeds recorded by Beurle(4). Applying the same principles to spoken English, with syllables being regarded as the individual stimuli, gives, as expected, a much higher speed.

2.2 At a higher level in the information processing system of our brains there is another restriction presented by the limited span of immediate memory. Miller(14) has described evidence on which he has based the hypothesis that our memory span for information is limited to about seven independent chunks of information, irrespective of the information content of each chunk. He has also suggested that the structure of language is evidence of the way we require to chunk or quantise information in terms of words, phrases, sentences, etc., and that this hierarchical structure indicates the way in which information has to be processed by our brains. Thus if the reading rate is slowed down by the response time of the ear or other factors the short term memory system is employed inefficiently and the beginning of a sentence or paragraph can be easily forgotten before the end is reached. The reading process therefore becomes extremely tedious and fatiguing.

2.3 Clearly, if a given passage of text is spoken, encoded into Morse or scanned by an optophone, the number of elements in the spoken version will be much fewer than in either Morse or the optophone code, although the amount of useful information conveyed by all three media will be the same. The reason for this difference stems from the fact that there is a high degree of redundancy in printed characters. This is illustrated by the familiar experiment of masking the lower half of a line of print whereupon it can usually still be read with very little difficulty. Much of the redundancy can be utilised by the eye and helps to minimise error, but, via direct encoding, it is not acceptable to the ear, which has a lower information capacity. When reading aloud the speaker carries out a complex process in which he removes the original redundancy of the print pattern and reproduces the essential information with redundancy of a different kind that the ear can utilise. The presence of redundancy in speech is indicated by its ability to resist many kinds of distortion and still remain intelligible.

2.4 One source of redundancy in print derives from the fact that letters are composed of continuous lines and, if given a partially masked character, the invisible portion may be guessed with small risk of error by inspecting the trajectories of lines at the boundary of the mask. In the case of the optophone, observation shows that the ear can make only very limited use of the sequential redundancy contained in successive chords and, if information is lost, it is much more difficult to guess correctly. The eye, having access to all parts of the two dimensional pattern simultaneously, is capable of utilising the redundancy in print; but the ear, receiving this same information serially, is not. This is not to suggest that the ear is unable to utilise sequential redundancy at all, for at higher levels, involving larger units of spoken text, sequential redundancy does play a part, but the rate at which the decisions have to be made is, by comparison, extremely low. The kind of redundant signal that the ear is accustomed to use is formed by duplicating the same information among several dimensions in the acoustic pattern. Thus, in the absence of information defining some signal property along any one dimension, there is a high probability that the same information can be obtained by the ear from other dimensions in the display. It will be clear from this discussion that the notion of direct coding must undergo some radical modifications aimed firstly at achieving a transformation of print which approaches more closely the principles involved in reading and, secondly, to ensure that the number of signal elements per word is reduced to syllable proportions.

2.5 Pollack(17)(18) has carried out experiments using sounds with different numbers of physical dimensions (e.g. frequency, loudness, rate of interruption, etc.). He has shown that multi-dimensional signals can convey more information per signal element than signals employing only one dimension. Thus increased dimensionality can be used not only to afford some duplication of information but also to carry more information per signal element. However, Pollack's subjects recorded their discriminations in their own time by a check list procedure. This meant that for the multi-dimensional stimuli containing more information, it was possible that a process of memorisation followed by serial identification of the dimensions made the recognition time proportionally longer. The importance of this point to the problem of specifying the output of a reading machine stimulated some work to determine whether multi-dimensional sounds did carry a penalty in recognition time.

2.6 In an outline of the psychological considerations that control the design of a reading machine output Studdert-Kennedy and Liberman(19) regard speech as permitting the highest attainable speeds of auditory communication and proceed to analyse the factors which give it this high efficiency. They state that there are good reasons for believing that there is more to the perception of speech than the fact that it is multidimensional and that discrete units of information (the syllables) fall within the resolution time of the ear. The authors describe

Voiced stops

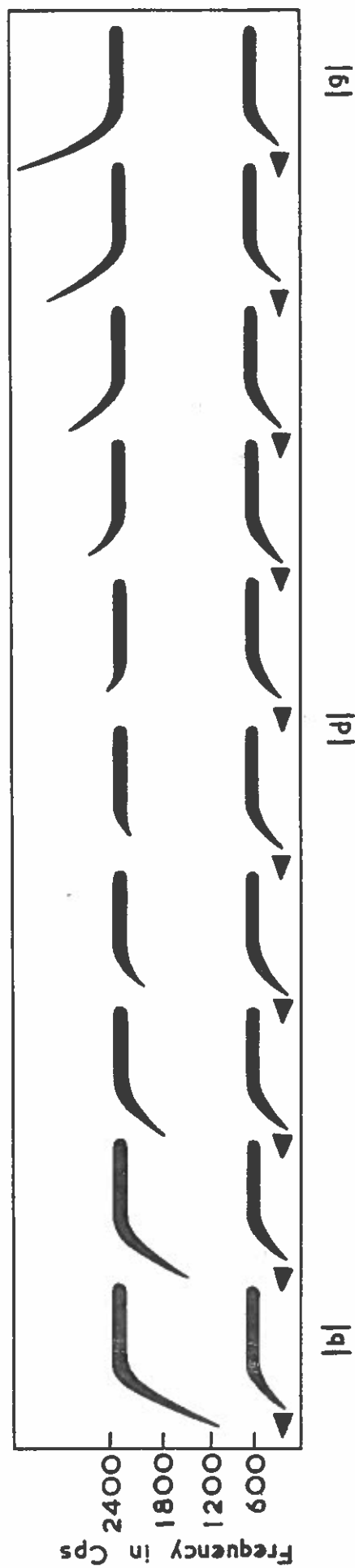


FIG. 1. SPECTROGRAPHIC PATTERNS THAT ILLUSTRATE THE TRANSITION CUES FOR VOICED STOPS

certain 'discrepancies' that occur in the perception of speech stimuli which they suggest may be explained by a theory of speech perception reported in their publications. This theory states that the perception of speech is tightly linked to the feedback from the listener's own speech mechanism and that we discriminate at least some speech sounds by monitoring the nervous activity necessary to imitate the incoming speech patterns. In a series of papers Liberman, Cooper, Fry, Eimas and others (12)(7)(6) have reported the effects of a gradual movement of the second formant transition in a series of synthesised phonemes, Fig.1. These sounds, which cover the consonants /b/ /d/ and /g/, are not heard as a series of gradually changing stimuli but as a series of identical /b/'s, followed by an abrupt change to a series of identical /d/'s, which again shifts abruptly to a series of /g/'s. Discrimination is shown to be sharp across the phoneme boundaries but to be very poor within phoneme categories. In contrast, corresponding discriminations of a similar number of stimuli spanning the three vowels /I/ /e/ and /æ/ do not show the same sharp categorisation and differential judgements (in which adjacent, or near adjacent, stimuli are presented in rapid succession) do produce precise results uniformly throughout the range. The theory suggests that the discontinuities that occur in the perception of the consonants reflect the different articulations that are necessary to produce them. Thus /b/ is produced by a movement of the lips and /d/ by a movement of the tongue. Unlike the vowels, there is no way in which a continuum of consonant sounds can be produced between /b/ and /d/. The articulatory movements are discrete and so are the perceptions.

2.7 If there is some feedback mechanism of discrimination which makes possible preferential responses to certain classes of speech sounds, then to achieve the best possible performance from a reading machine it is clearly desirable to use some form of analogue speech mechanism which produces these sounds. In view of the impact of the theory on our research it has been thought worthwhile to examine the evidence further.

2.8 The theory is intuitively plausible, but it has the unsatisfactory drawback that it is peculiarly difficult to design experiments that would really put it to the test. If examples of continuous perceptions in speech could be found, in situations where the sounds were produced by distinct articulations, only then could the theory be falsified. This appears to be the only direct method by which it could be tested satisfactorily. If the nervous feedback from the articulators were monitored by means of suitably positioned electrodes it seems possible that clear evidence could be gathered that would confirm or deny the theory, but in practice the techniques are difficult and, if no signals were detected, this result would provide insufficient justification for rejecting the theory. Indirect methods of examination involve setting up an alternative hypothesis and showing that the observations are consistent with the action of an entirely different mechanism. Thus we could postulate that the discrimination effects are learnt responses which could be achieved with non speech stimuli if sufficient time were devoted to training. However, the question of training time is crucial and will influence the results of any experiment to test this hypothesis, for it is unlikely that laboratory subjects could be trained for long enough to enable them to compete with speech discriminations at the same level of efficiency. Thus the results are unlikely to be conclusive.

There is one other approach to testing the feedback theory which does not suffer directly from the training difficulty. This approach involves an investigation of whether the critical differences in discrimination can be found for sequences of non mimicable sounds which are distinguished from one another by features analogous to those found in the synthesised speech stimuli. This point is examined in more detail in the following section.

2.9 There is an assumption in the methodology of the experiments with synthetic speech stimuli that both series of sounds are distributed along single dimensions in the perceived signal space and that there is not a prior reason why we should expect the results of absolute discriminations on /b/ /d/ or /g/ to be any different from those on /I/ /e/ or /ae/. In fact, from an inspection of the physical differences between the stimuli we certainly should be led to expect some differences in the responses. Firstly, the vowel stimuli have a single phase in which components which distinguish them from other vowels are sustained for 300 ms whilst the consonants contain two phases, the transient of the second formant and its subsequent stationary state during the following vowel. Although the total length of the consonant-like stimuli is also 300 ms, the portion of any one stimulus that distinguishes it from all others is only 75 ms in length. Taking into consideration the fact that the power in the second formant is 8 dB below that contained in the first, it is in consequence extremely difficult for the ear to detect rates of change of formant two differing by several hundred cycles/per/sec. This may explain the poor relative judgements that are made with stop consonant stimuli under ABX presentation procedures. Categorisation judgements on the vowel-sounds are therefore carried out under entirely different conditions to those for consonants. Genuine conditions for absolute judgements can prevail for vowels but classification of the consonants can be assisted by judging the trajectory of the 75 ms transient relative to the position of the following 225 ms vowel. Thus, the three states of the second formant transition, the descending, continuous and ascending phases, automatically define three categories for the observer. To summarize, an examination of the physical differences between the stimuli can provide an explanation of the inter-category similarities experienced with the stop consonants and can suggest some reasons for the sharpness of the category boundaries. However, these considerations are not sufficient in themselves to explain all the observations and it is necessary to make the additional proposal, that language training and experience help to sharpen perception in the regions of the phoneme boundaries. If the combination of physical factors and the effects of learning are indeed the determinants of the observations, then there is no reason why any non-speech sounds could not be utilised to form a communication medium as flexible as speech itself. Some preliminary experiments have been conducted to gather evidence bearing upon these conclusions.

5 DIMENSIONAL CODE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
DIRECTION	R	L	R	R	L	R	R	L	R	L	R	L	R	R	L	R	R	L	R	R	L	R	L	R	L	R
NOISE	+	-		+	-		+	-		+	-		+	-		+	-		+	-		+	-		+	-
MODULATION	+			-			+			-			+			-			+			-				
INTENSITY	H						L						H						L							
FREQUENCY	H (2000 cps)												L (480 cps)													

3 DIMENSIONAL CODE

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	
MODULATION	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z	x	y	z	x	y	
INTENSITY	H			M			L			H			M			L			H			M			L		
FREQUENCY	H (3000cps)									M (1200 cps)									L (480 cps)								

KEY		
R	RIGHT	
L	LEFT	
+	PRESENT	
-	ABSENT	
H	HIGH	
M	MEDIUM	
L	LOW	
x	ABSENT	
y	SINE	
z	SQUARE	

FIG. 2.

Chapter 3

Speed of Discrimination for Multi-dimensional Signals

3.0 Pollack and Ficks were able to show that, as the number of independent characteristics (dimensions) of a stimulus are increased, the amount of information transmitted per stimulus can also rise. But, as these writers themselves have pointed out, to be able to judge the relevance of this observation in understanding auditory communication, "it is necessary to see whether the transmission of information per unit time can likewise be appreciably increased". If increased discrimination speed could be shown to result from the use of a multi-dimensional display, this would help explain the superior performance of speech.

3.1 A direct test of whether multi-dimensionality affects the recognition time for a stimulus could be carried out by comparing a uni-dimensional code with a code having, say, five independent variables. Ideally the experiment should involve the discrimination of a sufficient number of stimuli so that the recognition time may be of the same order, or longer, than the reaction time of the response. The simplest variables to use in such an experiment would be frequency, intensity, noise, duration and modulation. It has been shown by Garner(8) and Pollack(17) and others that the number of absolute judgements that a subject is capable of making along any one of these dimensions is limited to about six. To furnish a sufficient number of distinguishable variables to satisfy the recognition time requirements, at least two dimensions would have to be used together. Here, frequency and intensity are the two most convenient dimensions. However, to achieve an adequate number of discriminable steps, the intensity dimensions requires a dynamic range of more than 95 dB, a requirement which would exclude the use of a taperecorder. But, because of its convenience and the need for reproducibility, the taperecorder is an essential tool and therefore in the experiment actually carried out compromises were made and the ideal choice of dimensions modified.

3.2 The dimensions that were finally chosen are shown in Fig.2. Some of these variables had been used earlier by Beddoes, Belyea and Gibson(3) to form the output of a proposed reading machine employing letter recognition. These authors carried out recognition tests on a number of dimensions, used together and in isolation, to establish relative independence and selected frequency, modulation and noise. To these we initially added time and intensity, but time was later abandoned because the recognition accuracy along this dimension was found to depend significantly upon the presence of noise. The variable finally chosen to replace time was direction of origin (i.e. left or right ear). All dimensions were tested singly and in pairs to ensure that the dimensions were equally discriminable.

The two codes comprised one of five dimensions (5D) and another of three (3D), constructed in such a way that both codes had approximately the same information content. The intensity levels used in the 5D code were 0 dB and -12 dB, and those for the 3D code were 0 dB, -10 dB and -20 dB; the maximum intensity level, 200 decibels, was adjusted arbitrarily by each subject. All signals were tape recorded giving a noise level at about -40 dB relative to the maximum signal intensity. The noise dimension was achieved by the addition of a separate noise source, raising the background noise level to -12 dB. Double side-band modulation giving either a square or sinusoidal envelope formed two states of the modulation dimension while the third was an unmodulated pure waveform.

3.3 Two groups of four subjects aged between 18 and 25 years were recruited and matched as closely as possible on the basis of academic ability. Each group began a course of training on letter recognition using one of the two codes. The

Percentage
Correct

Full Alphabet.

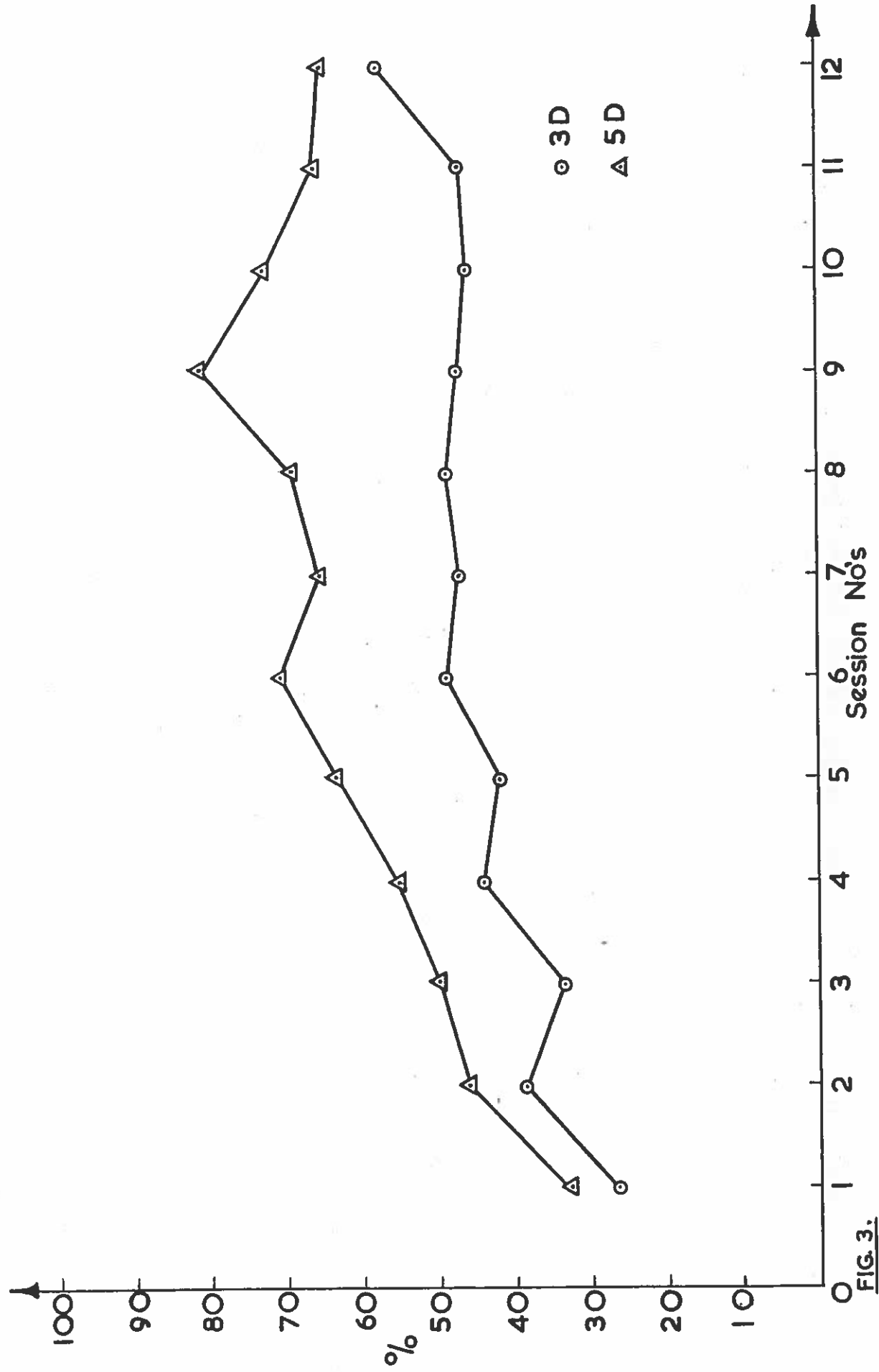


FIG. 3.

training method involved the division of the alphabet into three approximately equal sections which were learnt separately. The transition between one section and the next being made when the slowest member of the two groups had reached the 60 per cent correct level. When each section of the alphabet had been studied in this way the training on the complete alphabet was begun. The average performance of the two groups during training was monitored and is shown in Fig.3 for the nine sessions immediately prior to the three final test sessions (Nos. 10, 11 and 12). Each session lasted for between 15 and 20 minutes and was divided into four phases:

1. Earphone balance and adjustment
2. Short refresher course
3. Practice under test conditions
4. Recognition test of 40 symbols

The subjects attended regularly during a period of two months and during that time accumulated a total of five hours experience of the codes. The signals were all 0.6 secs in length and were followed by a three second interval during which the subject made his verbal response. Measurements were made of the time elapsing between the beginning of the signal and the beginning of the response for each of 120 symbols selected randomly from the full alphabet. All these results have been histogrammed in Fig.4. The distributions are not significantly changed if only the correct responses are considered. The complete set of response times were therefore pooled for each subject and group and the mean and variances calculated in Table 1.

TABLE 1

Group	Subject	Variance	Mean	Group Variance	Group Mean	t
3D	K.F.	0.34	2.24	0.48	2.34	4.9
	A.C.	0.25	1.95			
	M.W.	0.39	2.53			
	P.P.	0.61	2.63			
5D	R.C-H.	0.18	1.32	0.64	2.10	
	J.K.	0.22	2.16			
	R.H.	0.57	2.72			
	K.H.	0.61	2.18			

3.4 An information analysis (13) of the three test sessions revealed an average transmission of 3.30 bits/symbol from a possible 4.57 bits/symbol for the 5D code compared with 2.60 bits/symbol from a possible 4.62 bits/symbol for the 3D code. The information contributed by individual dimensions is shown in Table 2.

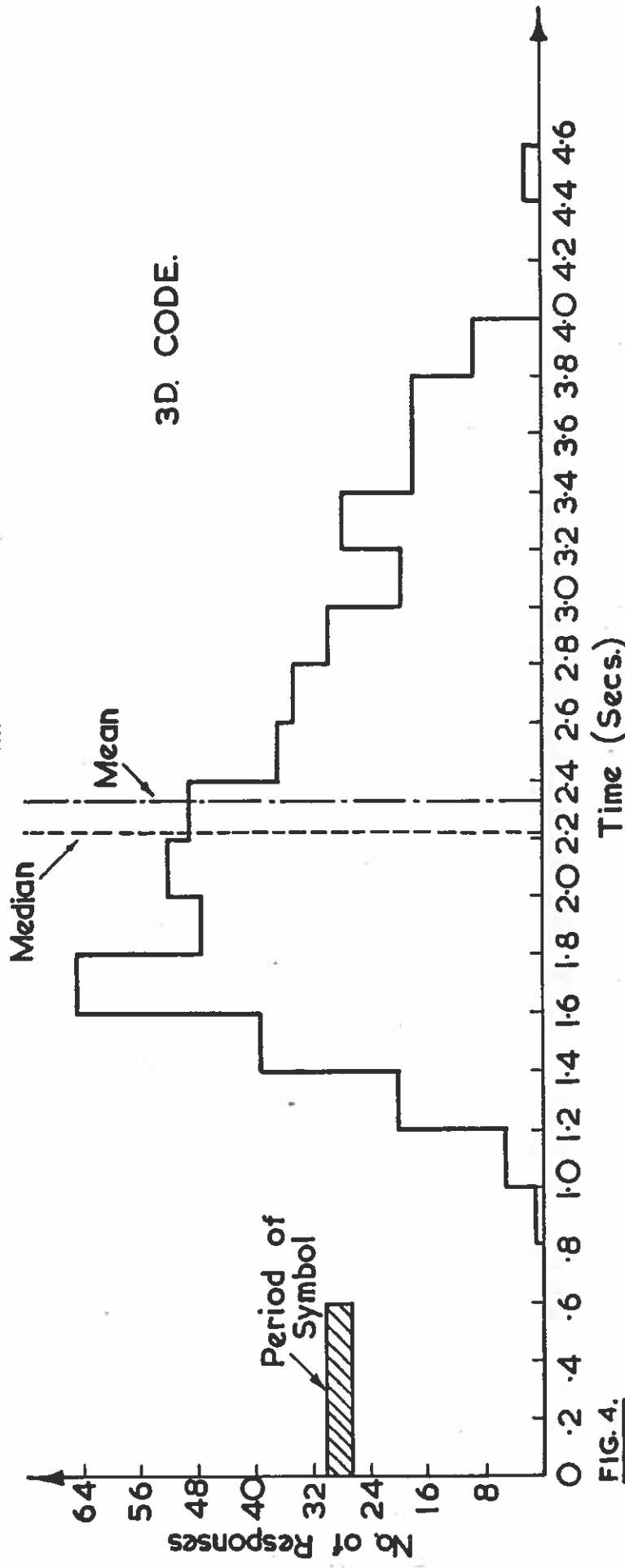
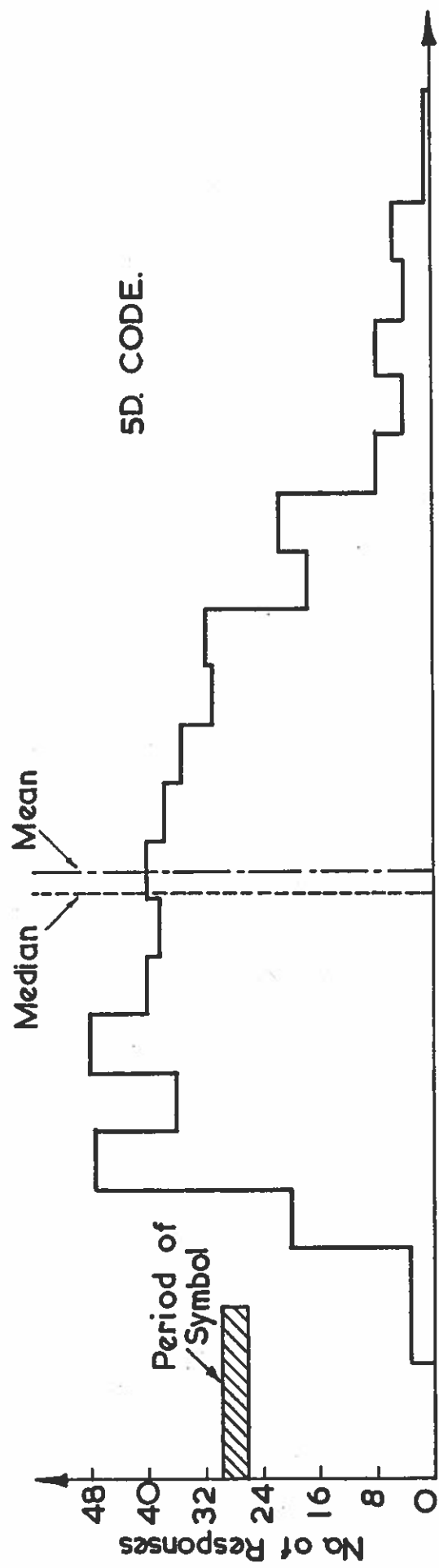


FIG. 4.

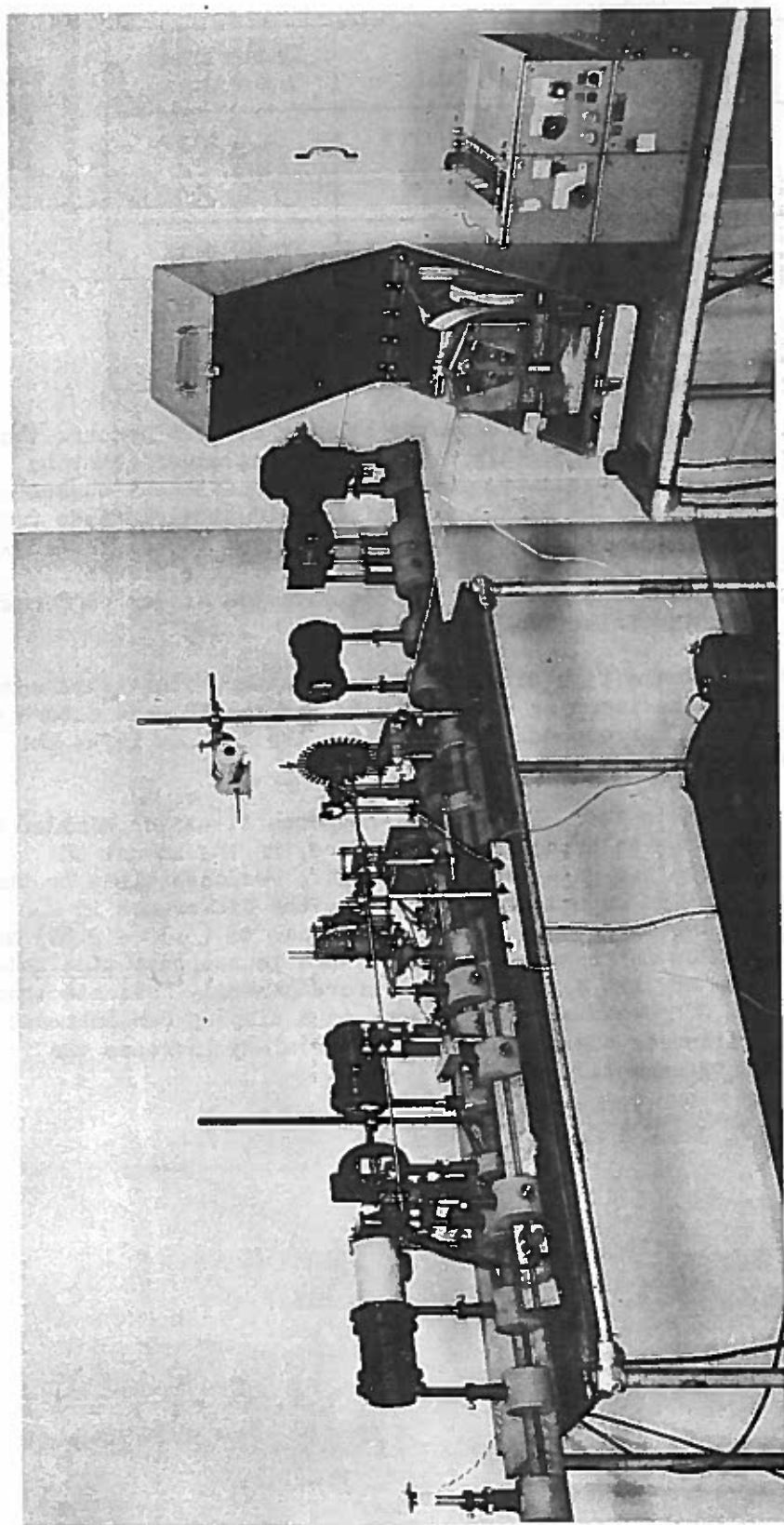
TABLE 2

Group	Dimensions	Input Information	Transmitted Information
3D	Frequency	1.58 bits	0.74 bits
	Intensity	1.58 "	0.43 "
	Modulation	1.58 "	0.81 "
5D	Frequency	1.00 bits	0.47 bits
	Intensity	1.00 "	0.55 "
	Modulation	1.00 "	0.55 "
	Noise	0.93 "	0.65 "
	Direction	0.97 "	0.61 "

The average response time to the 5D code was shorter by 0.24 seconds than the corresponding time for the 3D code, this difference being significant to better than the 0.001 level of probability. However, an insufficient number of subjects were available to permit the elimination of individual variations from the results and hence it is not possible to show directly that the differences between the performance of the two groups did not arise from this source. Nevertheless, there is no indication that the 5D code gave the slower performance expected from a check list discrimination procedure.

It should be noted that the five dimensional symbols were identified more accurately. The plateau in the learning curve, Fig.3, for the 5D code occurs at 70 per cent correct while the corresponding plateau for the 3D code is at the 49 per cent correct level.

Hick's(9) results indicate that, in a signal-response situation similar to our own, speed and accuracy can be exchanged. Therefore, if the amount of information transmitted by the two codes were made equal, response times to the 5D code could be decreased by an amount proportional to the difference in information transmitted by the two codes, i.e. proportional to (3.30 - 2.60) bits. This would result in a further increase of the difference in response time between the two codes. The results of this experiment therefore strongly indicate that any increase in the number of perceivable dimensions in a display can increase the rate of transmission for discrete signals and almost certainly increase the continuous transmission rate as well.



SCANNING MECHANISM TO PRODUCE PAPER TAPE INPUT MATERIAL
FOR THE PROCESSING PROGRAM.

FIG. 5.

duplex

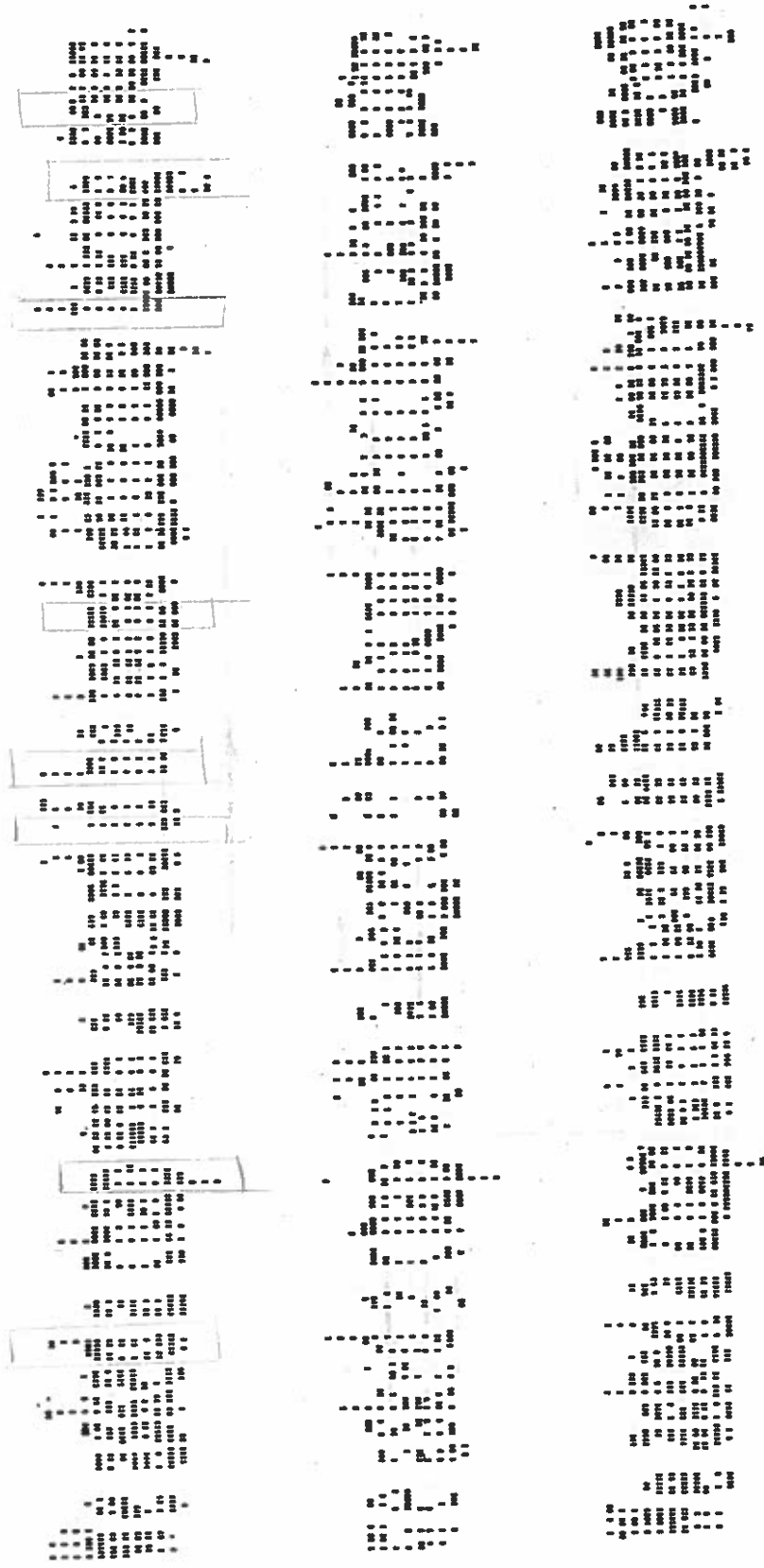


FIG. 6.

Chapter 4

The Simulation of different output Displays

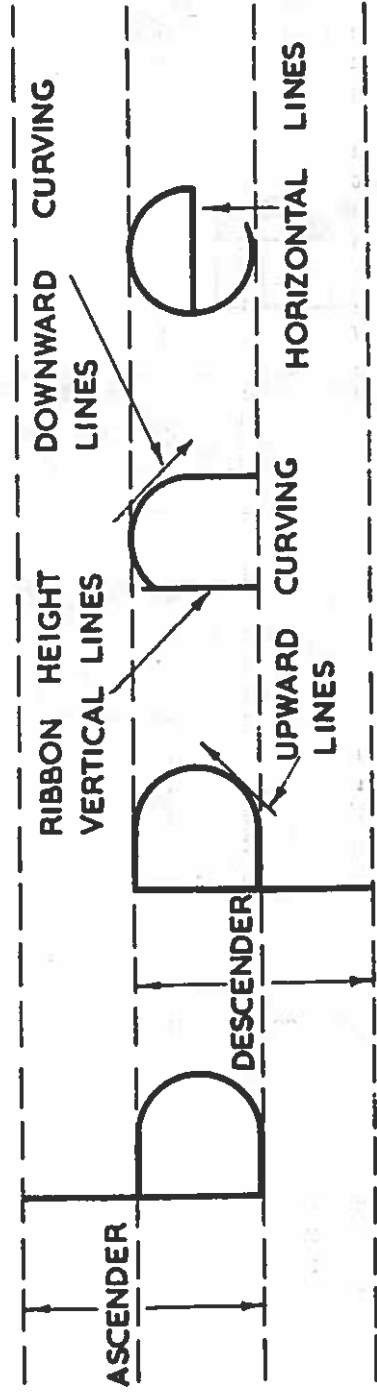
4.0 Before proceeding further, it was necessary to construct a flexible system which would provide a range of different outputs for experimental use. The flexibility arose from the fact that the range of outputs could be controlled by two kinds of input signals; the first derived from digital information gathered by a single row of photocells scanning the characters and the second from a more complex system which endeavoured to carry out feature detection on curves and lines and so reduce this source of redundancy. Many different interconnections of these input and output units were possible but only those connections which provided outputs relevant to the questions under study were used. A considerable portion of the system was built from laboratory optical and electronic equipment and the remainder simulated on a digital computer. The computer program was written with the aim of achieving certain specific objectives in the simplest and most convenient way, and without regard to exploring any particular design philosophy. None of the equipment was therefore constructed with an eye on economy. In essence, the strategy adopted was to build an apparatus which, in one form, produced an output conforming in a number of ad hoc ways to the principles demanded by certain conceptual models, and to measure the extent to which the rejection of one or more of these principles (by using other input output relationships) affected the overall performance of experimental subjects.

4.1 The basic input to the system, from the printed page, was provided in the early stages by an optical scanning device shown in Fig.5 and described in Appendix B. Fig.6 shows the output from this apparatus recorded on three different occasions. It was found that, to achieve consistent results, the mechanical alignment had to be maintained within close limits. The plan to use this equipment to generate new data when required was abandoned as a direct result of this experience, but the clearest scan patterns of individual letters of the alphabet were salvaged and used as a standard source from which all subsequent digitised text material was assembled by hand. These patterns were then either degraded into a six bits per scan form closely equivalent to the optophone, or were processed directly by a simple feature analysis program in a digital computer. The object of this analysis was to effect a reduction of the amount of binary data in the source patterns whilst essentially retaining the important textual information.

On average each character was represented on a 12 x 5 bit matrix; for the letter (m), the widest character, the size of the matrix was 12 x 9 bits. Connected text was assembled with one scan of zero digits inserted to separate each character from its neighbour and five blank scans were used to separate words.

4.2 The feature analysis program was a highly simplified version of the technique described by Uyehara(20) and functioned in three stages. The first stage consisted of an examination of the first column of digits and identification, by comparison with a mask, of ascenders, ribbon height verticals and descenders, Fig.7. Stage two involved the storing of this column and the comparison of the next scan line or column with the store contents to determine whether there were digits present in identical positions in both columns - indicating the presence of a horizontal line - or whether digits were displaced above or below one another - indicating upward or downward curving lines. These stages were repeated with successive scans. A total of six features were therefore selected and each identification recorded individually by a token placed in the respective location. In Fig.8 these locations are set out as the columns of matrices produced by each of the lower case letters (a - v) from the standard alphabet. The order in which the columns are set out from left to right is 1. Ascenders, 2. Descenders, 3. Ribbon Height Verticals, 4. Curvature Downward, 5. Curvature Upward, 6. Horizontal Continuity.

SIMPLE SCANNING PROCESS FOR IDENTIFYING LOCAL FEATURES IN PRINT



CHARACTERS ARE SCANNED
VERTICALLY AND STORED
DIGITALLY AS SHOWN FOR
THE LETTER b.

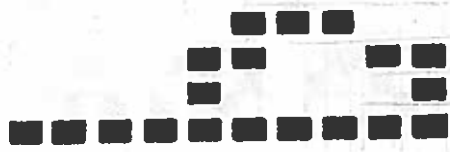
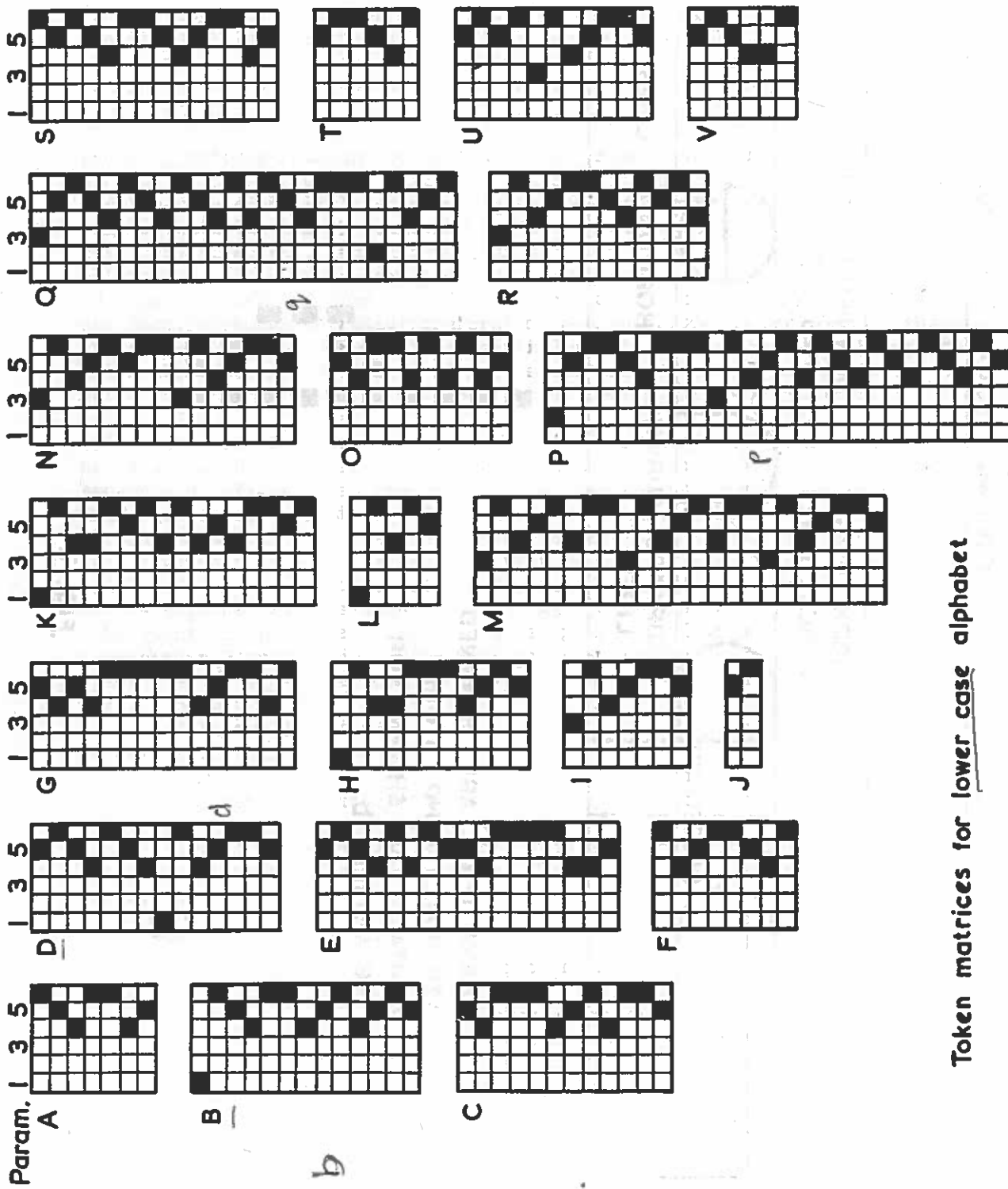


FIG. 7.



Token matrices for lower case alphabet

FIG. 8.

4.3 At the present stage of development, if characters from a different font had been used or they had not been correctly aligned, the detailed structure of each of the matrices would change. However, for the case where they originate from different digital copies of the same character, broad similarities between the matrices would still remain. A method was sought that would succeed in filtering out the small discrepancies in these matrices and allow only the gross differences between different letters to emerge. The technique chosen to achieve this involved the blurring of each non-zero binary token by a Gaussian function (the equivalent of passing the output of each channel through a low pass filter), thereby producing six continuous functions which were punched out in multiplex form on paper tape.

There are a number of ways in which the filtering effect could have been achieved. The final choice however was not entirely arbitrary and can be justified partly in terms of expediency and partly in an endeavour to make the process parallel some of the activities involved in reading and speaking. Thus the effect of the Gaussian functions was to perform quasi-integration of the token matrices and, by presenting information the form of continuous waveforms, to provide a method of achieving the coalescence of groups of letters into word units. An analogous situation exists in sighted reading when under normal conditions the reader is able to recognise a word by its overall shape rather than by the detailed structure of the letters. When context fails to supplement this loss of letter information the reader then reverts to a study of individual letters. The corresponding process in the simulation is controlled by varying the standard deviation of the Gaussian functions, although this was never explored. As the functions become narrower, it is to be expected that more letter detail can be made available, but with a consequent loss in reading speed due to the ear's limiting reception rate for discrete stimuli.

4.4 It will be seen that the feature analysis program reduces the amount of data available at the input by removing some of the redundant information. Roughly how much data is removed can be shown by calculation based upon a few simplifying assumptions. If the original input patterns are presented by an optophone-like harmonic display, which has been thoroughly learnt, then it would be reasonable to assume that the reader would know intuitively the likelihood of occurrence of each of the twelve elements of the scan. Hence the reader can be regarded as an optimum decoding device and the information contained in rows of digits along a length of text can be calculated from the Shannon Wiener equation $H = - \sum p_i \log_2 p_i$. We must also assume that the interdependence between rows (i.e. shared information) is not perceived by the listener and that they can therefore be analysed independently.

For the selected characters which form the standard alphabet, the analysis on a row by row basis, using letter frequencies found in English, gives an average information per character of 37 bits. The corresponding calculation carried out on the token matrices derived from the standard alphabet requires the reasonable assumption that the columns representing the parameters are perceptually independent. On this basis the average information per character is approximately 22 bits and the reduction factor is therefore about 1.7. The ideal factor for complete recognition by the processing device is $37/\log_2 26 = 7.9$ which can be approached more and more closely as the feature analysis is refined and made more efficient. The greatest inefficiency arises from the existing technique for detecting curves and horizontal continuity which uses only two consecutive scans. An increase in the horizontal range over which this detection logic functions would be one way of improving efficiency but the complexity of the logic would increase and make any attempt at an economic hardware realisation more difficult.

4.5 The paper tapes carrying the feature waveforms were converted into six simultaneously available analogue voltage functions by means of an electronic de-multiplexing device and digital to analogue converter. A conversion rate of

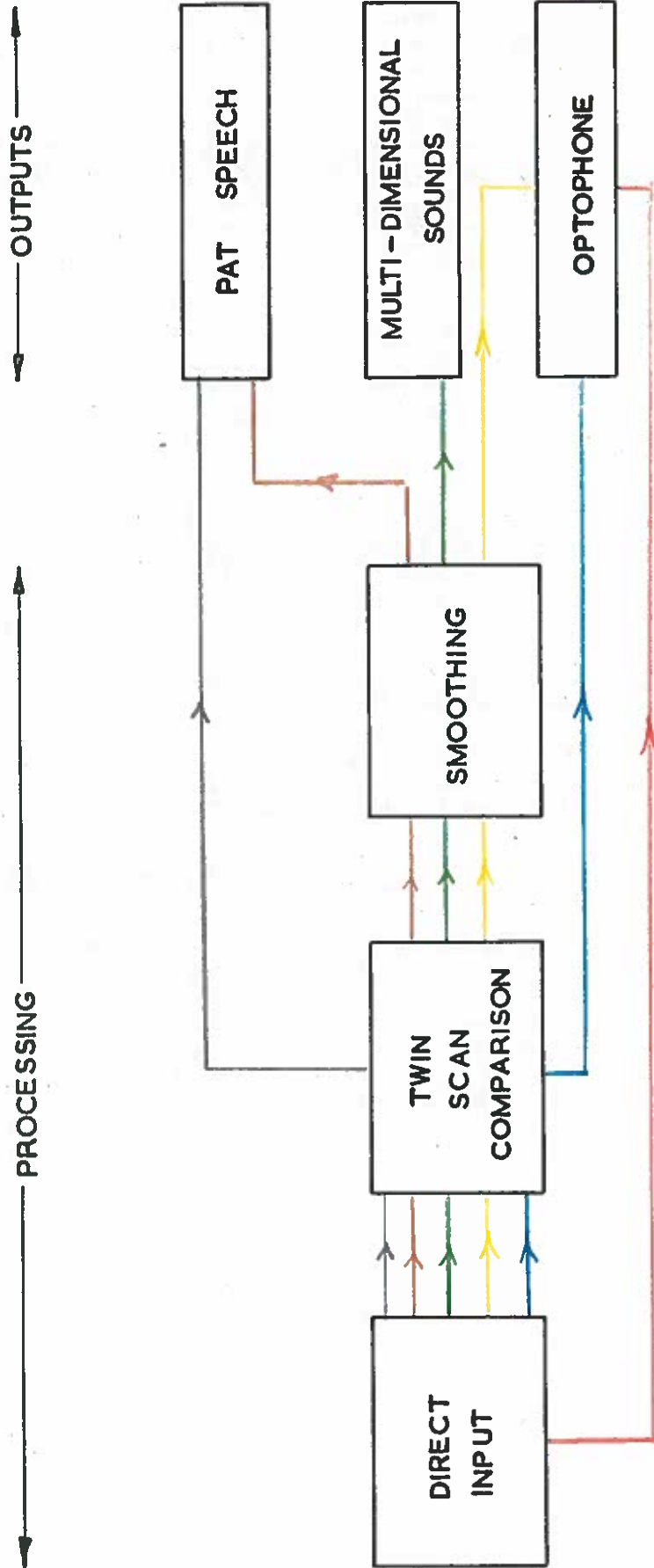


FIG. 9.

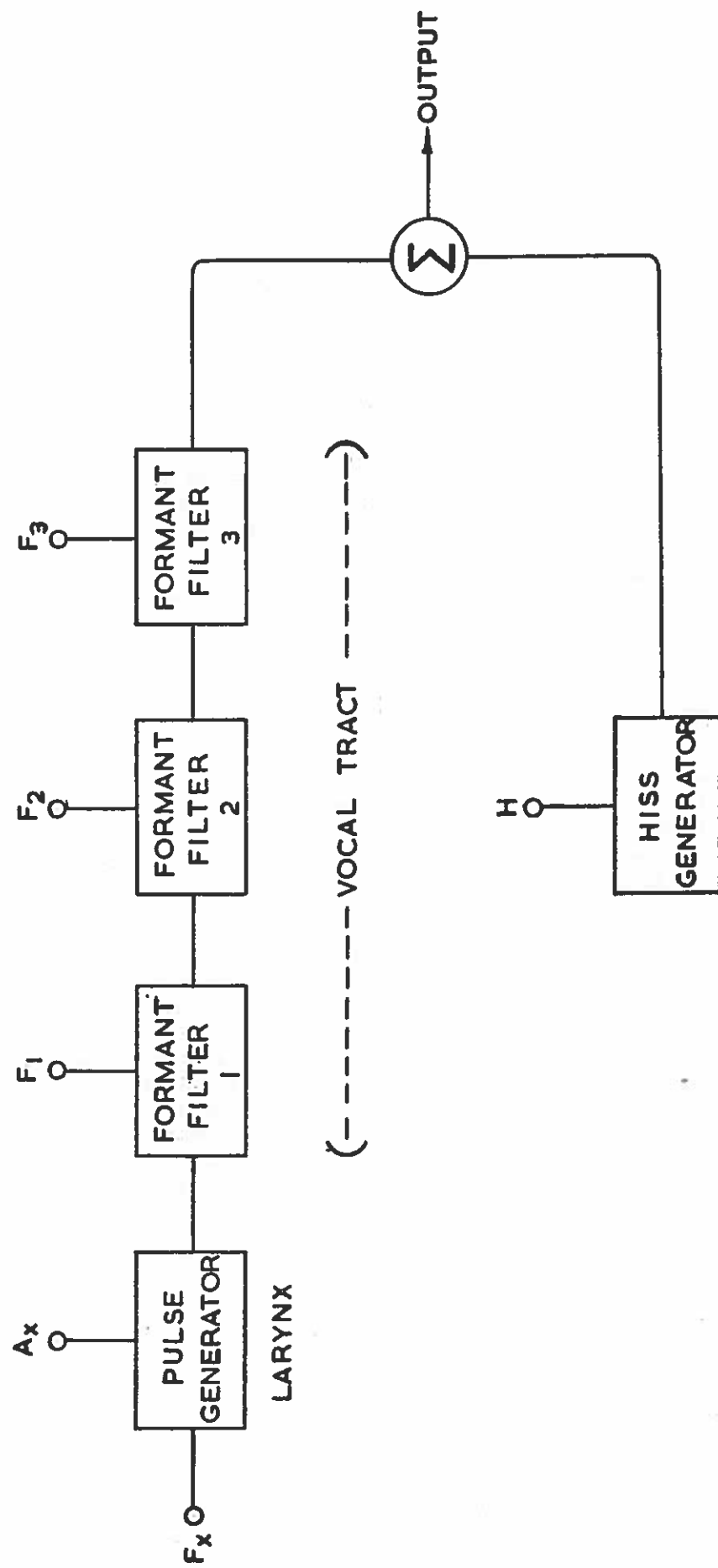


FIG.10. A BLOCK DIAGRAM OF THE 'PARAMETRIC ARTIFICIAL TALKER' (PAT)

60 words per minute was usually chosen which meant that the Gaussian functions had an effective standard deviation of 0.18 secs. The set of control functions could be connected to one of three outputs; an optophone display, a multi-dimensional sound generator or a speech synthesisor. Provision was made for some by-pass connections. For example, a six bit per scan degraded version of the input patterns, punched on paper tape, could be used to trigger six tuned oscillators thereby producing a conventional optophone output similar to that produced by the Battellé reader. Fig.9 illustrates some of the available connections between the major components.

4.6 The optophone oscillators were tuned to frequencies of 1,000 cps, 768 cps, 640 cps, 576 cps, 512 cps and 384 cps, and could be switched on abruptly by signals derived directly from the input patterns or could be controlled by the continuous functions. These six independent functions could be made to continuously vary individual oscillator outputs over a range of 40 dB by means of electronic volume controls.

4.7 The multi-dimensional sound generator was evolved from the equipment used in the experiment described in section 3.2. Only four of the original dimensions were used however; noise, square wave modulation, frequency and intensity. In addition the generator included a circuit producing a click. The frequency of the output was switched between 480 cps and 2000 cps by means of a trigger circuit with the threshold set at the 60% level and driven by the control functions. The click generator was controlled in a similar manner. The noise, depth of modulation and overall output intensity were all controlled by the waveforms via electronic volume controls. The feature waveforms were associated with the components of the multi-dimensional generator in the following manner:

- | | | |
|----|-------------------------|-----------------------|
| 1. | Ascenders | - Click |
| 2. | Descenders | |
| 3. | Ribbon Height Verticals | - Noise intensity |
| 4. | Curvature Downward | - Overall intensity |
| 5. | Curvature Upward | - Depth of Modulation |
| 6. | Horizontal Continuity | - Frequency |

4.8 The speech synthesisor used in the experiment was an early version of the Parametric Artificial Talking (PAT) device designed by Lawrence(11)(2). The device is an electrical analogue of the human vocal mechanism and comprises components which simulate the action of the larynx and the cavities of the vocal tract, Fig.10. The larynx wave form generator produces a pulsed output whose frequency and loudness can be controlled by applied parameter voltages. This signal is fed through three filters connected in series whose resonant frequencies are also controlled by external voltages. The filters simulate the action of the vocal cavities and impress the characteristic formant patterns of speech on the emerging waveform. A sixth parameter controls the output of a noise generator which is mixed with the filter output, amplified and fed to a loudspeaker. The six controls required by this system correspond to the six sources of information arising from the computer program. With the exception of the experiment described in section 6.3 the following features were employed to control the various components of the synthesisor

- | | | |
|----|-------------------------|--------------------|
| 1. | Ascenders | - Noise |
| 2. | Ribbon Height Verticals | - Formant 3 |
| 3. | Curvature Downward | - Formant 2 |
| 4. | Curvature Upward | - Formant 1 |
| 5. | Horizontal Continuity | - Larynx Frequency |
| 6. | _____ | - Larynx Amplitude |

A trail passage of text: The quick brown fox jumps over the lazy dog was processed by the computer program and the waveforms compared with those required

to generate intelligible speech. Allocation of the control waveforms to components of the PAT machine was decided on the basis of achieving the closest possible match with the speech parameters. The sixth control took the form of a rectangular waveform switching on when parameter 5 exceeded zero volts and off when the parameter fell once again to zero.

Details of some aspects of circuit design are given in Appendix C for the more important apparatus used in this work.

Chapter 5

The Methods used for Evaluating Different Auditory Outputs

5.0 Two basic methods of training subjects were used. The first aimed at monitoring the rate of learning in the very early stages under forced choice conditions and the second method measured performance after several hours of self-paced tuition.

To permit a comparison of our results with those of two earlier workers ((Cooper(5) and Abma(1)) a method derived from that described originally by Cooper was used. Thus, what was termed the "Haskins Method" utilised a vocabulary of eight words which were transformed into different sound patterns. Tape recordings which displayed each set of these sounds were made in two sections. Section 1 was composed of an introduction to the eight word sounds in which each sound was played twice. Section 2 was formed by the test passages in which eight or sixteen series, each of 24 sounds, were played with a break of 45 seconds between each series. Within each series each sound appeared a total of three times with the order of appearance randomised.

The presentation sequence for each word-sound followed a procedure in which the sound was played once, there was then a pause of three seconds whilst the subject recorded his identification, whereupon the sound was played again and immediately followed by the correct answer. A further four seconds elapsed and the cycle began again with a new word-sound. Each subject learnt the response list of words before the experiment began. Between two and three minutes were set aside for this purpose before Section 1 of the tape recording was begun. Having heard each word-sound played twice the subjects proceeded to the test Section. The word identified with each sound was written down by the subject on a sheet of paper. Each subject's progress was measured in terms of the number of sounds correctly identified in each series of 24. From this data learning curves could be compared with one another.

5.1 The long term self-paced training programme was called the "NPL Method". Recordings consisted of a short introduction termed Section 1 in which each letter or word-sound, taken in turn, was played five times; Section 2 in which the eight sounds were played through three times; Section 3 in which pairs of sounds which promised to be difficult to distinguish were played alternately to assist discrimination and Section 4 in which sounds were played in random order and their identity announced after each presentation. Following the training sections were 18 practice passages each of twenty sound symbols and the final tests of 240 symbols. Correct answers were provided for all the practice sequences and subjects were thus able to monitor their own progress.

The length of the training period varied between three and six hours but very little improvement in performance was noted in the last three hours; in fact, most subjects achieved a 50% correct level of accuracy within the first hour. The final tests provided the data for overall analysis and consisted of six sequences of 40 symbols. Two sequences were played at 15 symbols per minute, a further two sequences at 20 symbols per minute and the remainder at 25 symbols per minute. With the exception of one experiment, all the subjects trained by this method were provided with individual tape recorders in a language laboratory. They could therefore test themselves on the material provided and when necessary rewind the tape back to the introductory sections and investigate the source of their errors. The responses to the final tests were written down and instructions were given not to stop the tape recorders during this period.

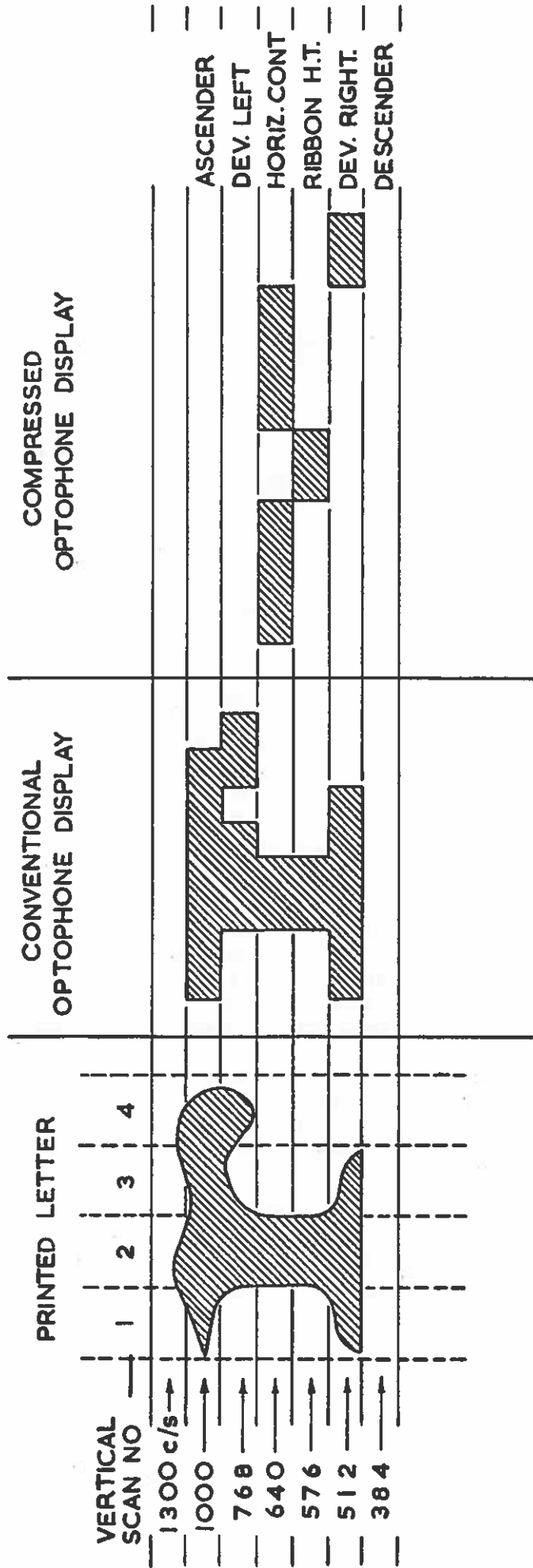


FIG. 11.

Chapter 6

The Experiments and their Analysis

6.0 The experiments described in the following Chapter were each performed with the object of gathering evidence bearing upon what advantages if any accrue from the use of data compression or a speech-like output in a reading machine. Even if the answers promised to be clear cut at the start of an experiment, at the end it was sometimes found that the results could not be interpreted in quite the way expected. An example of just such an experiment was the attempt to measure the effects of removing some sequential redundancy from the optophone, whilst still maintaining the same dimensionality of the display. Two sound displays were therefore generated using the units linked by red and blue lines in Fig. 9.

The system connected by a red line produced a "Conventional optophone" output similar to the Battelle reader. The connections shown in blue produced a system called the "Compressed optophone" which operated upon the degraded 6 bit scan and identified Ascenders, Descenders, Ribbon Height Verticals, etc. and switched on oscillators associated with a particular feature as follows.

TABLE 3

1000 cps	Ascenders
768 "	Curvature upward
640 "	Horizontal lines
576 "	Ribbon height verticals
512 "	Curvature downward
384 "	Descenders

Fig. 11 shows the results of this processing procedure functioning with the lower case letter (r). Thus the "Compressed optophone" display is produced by the feature detection program already described which controls the different frequency channels through a paper tape reader. There is no delay or integration process involved and, therefore, the serial form of the information emerging from the computer is preserved in the output. Eight lower case letters were chosen as the source material; f, i, k, j, p, q, r, z.

6.1 Using the 'N.P.L. Method', two teams of five paid subjects, aged between 18 and 20 years, were engaged and trained for a period of $3\frac{1}{4}$ hours on one of the two groups of sounds. When the tests were completed the two teams were exchanged and the training and testing procedure repeated.

Confusion matrices were constructed from the final test data for the pooled results of the two teams and a multi-variate analysis carried out - see Appendix (D). The results of the analysis and a statistical test of significance are shown in Tables 4 and 5.

TABLE 4

Output	Information Transmitted	Max. Possible Information Transfer
Conventional	1.92 bits/symbol	2.99 bits/symbol
Compressed	2.46 bits/symbol	2.99 bits/symbol

TABLE 5

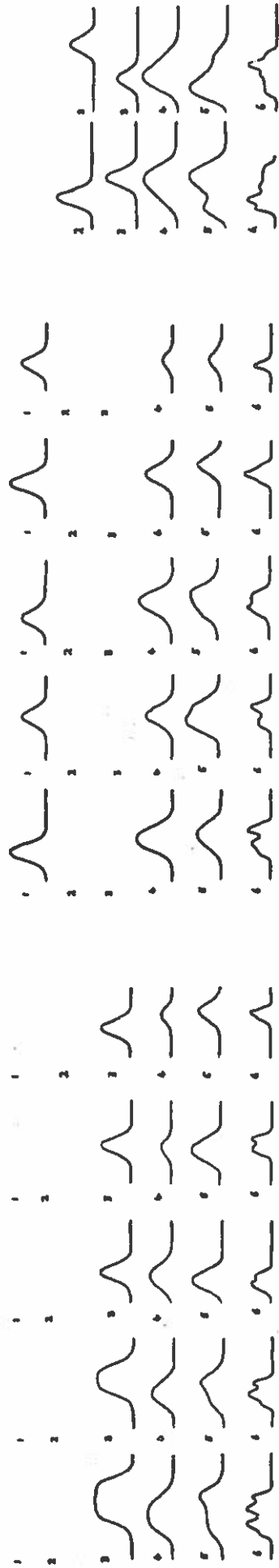
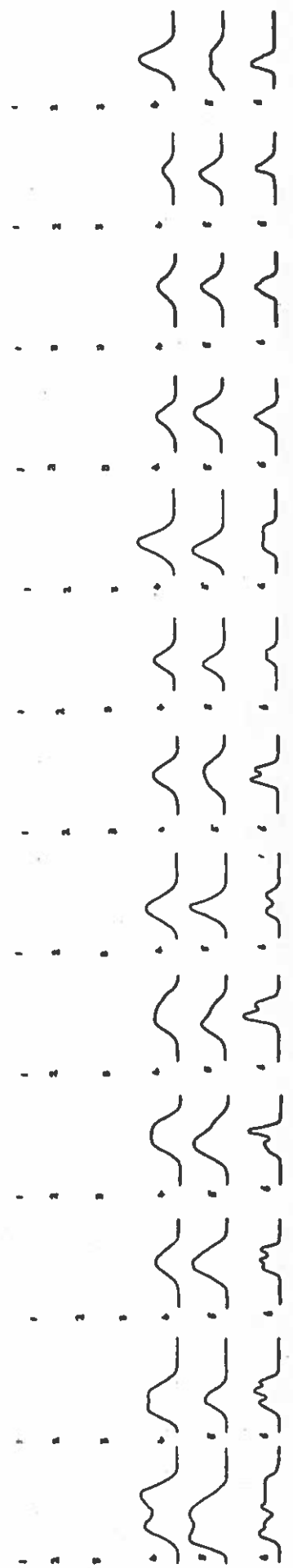
No. of degrees of freedom = 9
The pooled results of 10 subjects

Output	Mean of No. Correct (Max. 40)	Variance	"t"	Significance
Conventional	30.70	9.12	3.03	P = 0.015
Compressed	36.47	5.00		

Table 5 shows that the performance of the compressed optophone system was significantly superior to the conventional optophone. This occurred despite the fact that the variances in the presentation length among the letters was larger in the case of the conventional optophone and therefore provided additional cues for identification.

6.2 The results of the experiment appear to point to the view that reducing the redundancy present in the display can produce a significant improvement in response accuracy. However, this depends upon a number of assumptions. Taking, as an example, the general case of a transmitter, altogether there are three variables involved; the dimensionality of the output, the rate at which information is being transmitted and the amount of redundancy present. All these variables are inter-dependent, for, if the dimensionality is held constant and redundancy is reduced, the information transmission rate must automatically rise. Thus, the deliberate alteration of any one variable results in the actual alteration of at least two. The observations of this experiment are being made on the transmitter and receiver as a single entity and the interpretation of these results in terms of the response of the receiver to one particular altered parameter must assume that the other factors which are also varied have no effect. In fact, by examining the subjects' introspective observations, this assumption was shown to be not-valid. Thus changes in all three variables discussed above could have been perceived and it is not possible to draw the unequivocal conclusion that the experiment demonstrates the need for the removal of some of the geometric redundancy of print. Without much more knowledge about precisely what are the perceived dimensions in an acoustic signal there is little prospect of avoiding this dilemma.

6.3 The first experiment using the speech-like output, produced by the components connected by a brown line (Fig. 9), explored the differences in recognition accuracy between letter symbols produced by PAT and those from the conventional optophone. Further, the experiment sought to discover whether there was any correlation between visual discriminability at the parameter waveform level,




and auditory discrimination tests carried out on the sounds produced. To this end the parameter waveforms representing each of the 26 lower case letters were drawn on separate cards and a number of subjects were asked "to classify them into groups having broadly similar characteristics". All subjects chose the most obvious step, namely to divide the cards in such a way that each group contained non-zero values in common parametric positions (see Fig. 12). The subjects were then asked to take each group in turn and order the cards, placing the card with the most striking features on the left and ranging the remainder linearly so that the differences between adjacent cards were minimal. The results were remarkably consistent. The most frequently adopted ordering is shown in Fig. 12, and can be identified by reference to Table 6.

TABLE 6

					j							
m	n	r	u	i	k	d	b	h	l	p	q	
w	z	s	e	g	x	c	v	y	o	a	f	t

On the basis of this ordering two groups of eight letters were selected. The first (Group 1) consisted of the eight adjacent characters, s, e, g, x, c, v, y, o from the bottom row, which were termed the similar letters. The second (Group 2) was composed of eight characters selected from the extremes of each of the principal groupings, eg. j, m, i, k, l, p, w, t.

Recordings of the speech sounds were made on this occasion at the Phonetics Department of Edinburgh University and the connections to the control parameters differed slightly from that given in Section 4.8. The six control wave forms were connected to the synthesisor as follows:

1. Ascenders - Hiss 1 (S)
2. Descenders - Hiss 2 (s)
3. Ribbon Height Verticals - Formant 3
4. Curvature Downward - Formant 2
5. Curvature Upward - Formant 1
6. Horizontal Continuity - Larynx Frequency
7.  - Larynx Amplitude

The seventh control was a rectangular waveform switching on when parameter six exceeded zero volts and off when the parameter fell once again to zero. Two recordings were made with these sounds, which taught an association between them and the letters A - H or the numbers 1 - 8. A third group of lower case letters (b, f, g, q, u, v, w, z) which had been selected at random were recorded with the conventional optophone system, and a training tape constructed which related these sounds (Group 3) to the letters S - Z.

6.4 Twelve paid subjects were divided into six teams and were trained by the N.P.L. Method on all three groups of sounds. Order effects were balanced by the use of a 3 x 3 latin square procedure. When all the data had been collected the

confusion matrices for the groups were constructed and the means and variances calculated for the pooled results from the final tests.

The results of a multivariate informational analysis of the confusion matrices are shown in Table 7.

TABLE 7

Group	Information Transmitted	Max. possible Information Transfer
1	2.61 bits/symbol	2.99 bits/symbol
2	2.57 bits/symbol	2.99 bits/symbol
3	1.83 bits/symbol	2.99 bits/symbol

The information transfer achieved with the two speech-like outputs is clearly shown to be superior to that achieved with the optophone system. The statistical significance of this result is illustrated by Table 8.

TABLE 8

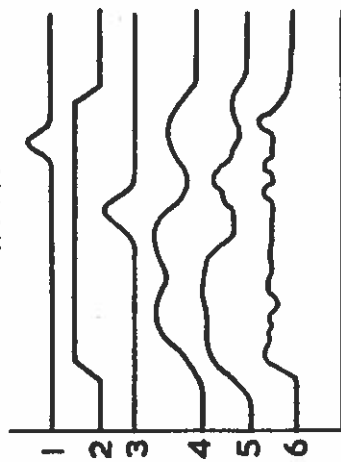
No. of degrees of freedom = 11
The pooled results of 12 subjects

Group	Mean of No. correct (Max. 40)	Variance	"t"	Significance
1	37.6	11.3	between 1 & 2 = 1.76	P = 0.1
2	36.3	18.5	between 2 & 3 = 4.58	P 0.001
3	28.6	42.8	between 1 & 3 = 5.63	P 0.001

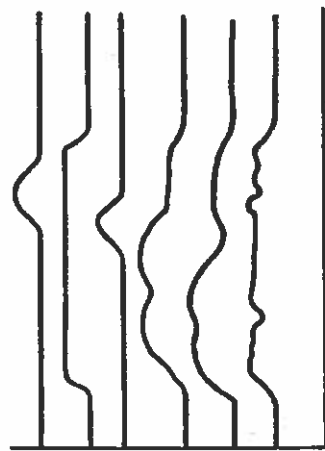
The other fact that emerges is that the performance on the dissimilar and similar section of control waveforms is not significantly different. However, a closer examination of the confusion matrices, shown in Appendix (D), reveals that 57% of the total errors occurring for the dissimilar Group 2 sounds is contributed by a confusion between only one pair of symbols A & E alias (j) and (l). If the two highest sources of error for the sounds of Groups 1 and 2 are neglected (13% in the case of Group 1), the dissimilar sounds are found to be superior in performance to the similar sounds. The reason for the confusion arises because the power of the noise component, indicating the presence of an ascender for (l), is very low. Furthermore, the variation in the Formant 2 parameter for (l), indicating curvature, is also small and, although differences between the sounds are obvious if they are rapidly alternated, absolute discrimination of the differences is extremely difficult.

6.5 To study the effects of random noise and a restricted band-width on the discrimination accuracy, the final test material was readministered to the

WITH



WILL



1. Noise

2. Larynx amplitude

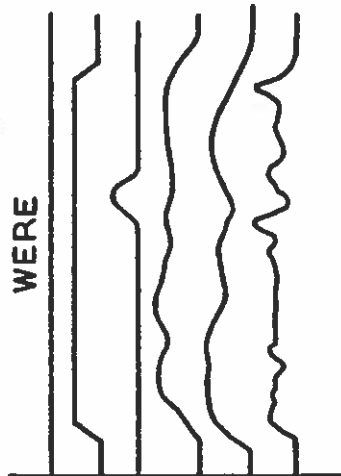
3. Formant 3

4. Formant 2

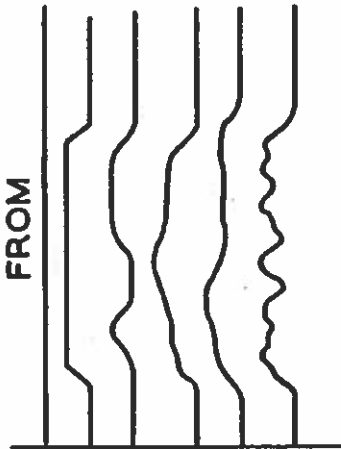
5. Formant 1

6. Larynx frequency

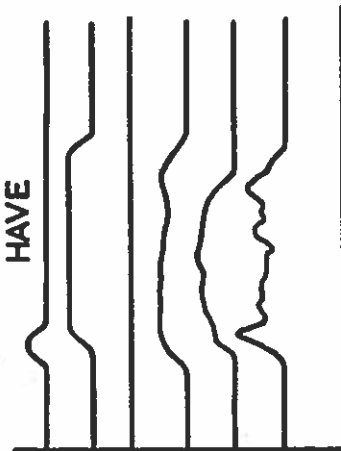
WERE



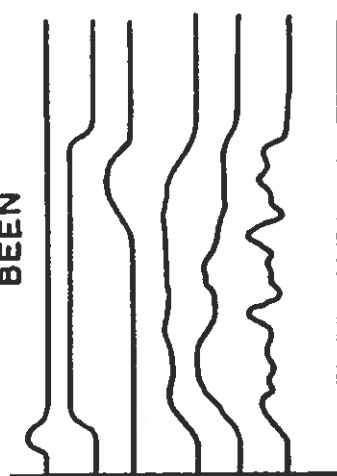
FROM



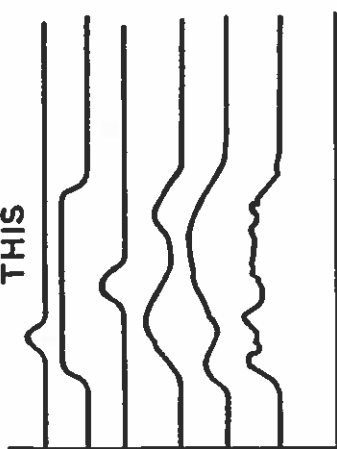
HAVE



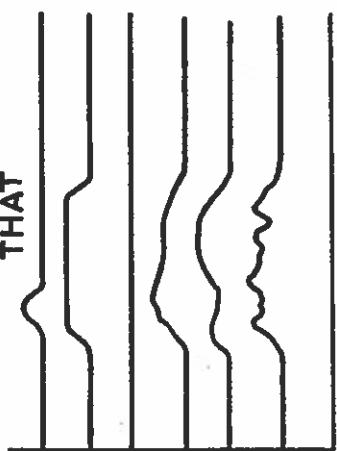
BEEN



THIS



THAT



subjects through poor reproducing equipment having a pass band extending from 100 cps to 1.8 Kc/s and reducing by 12 db per decade at higher frequencies. The noise level was in the region of 36 db with respect to the peak signal intensity. Table 9 shows the scores obtained on this occasion expressed as a percentage of the score obtained in the earlier test.

TABLE 9

Group	Score - Second Attempt
1	92%
2	90%
3	64%

6.6 These results show that the speech-like output has two principal advantages; firstly, it can produce sounds which are more easily discriminated than those from the optophone; secondly, the speech output is far less vulnerable to the effects of poor reproduction equipment and noisy environment. The question arises whether these properties are unique to speech or whether they are shared by all multi-dimensional signals. In the following experiment the rate of learning for multi-dimensional audio signals was compared with speech-like sounds and the results of some earlier work.

6.7 In "Research on Reading Machines for the Blind" (5), the final report published by F. S. Cooper and P. A. Zahl in 1947, an experiment was described which furnished learning curves for a number of projected reading machine outputs. The principal point that emerged from the Haskins experiment was that a speech-like code derived from spoken phonemes produced a performance superior to a wide range of direct coding reading aids. A basic vocabulary of eight words, identical with that used by Cooper, was used here as the generating source for three audio outputs. These were sounds produced by the systems connected by green and brown lines in Fig. 9 and termed the MDO and PAT sounds, and in addition a natural speech code (called Wuhzi) generated by a human speaker according to the rules described in Haskins report. Thus the speech code was derived from the word list by substituting vowels by other vowels and stop consonants by other stop consonants, etc. The vocabulary and code words are listed in Table 10.

TABLE 10

With	Yekw
Will	Yemm
Were	Yini
From	Snal
Been	Jiir
Have	Wozi
This	Kwef
That	Kwok

The waveforms produced by the feature analysis program are shown in Fig. 13.

6.8 Sixteen volunteer subjects, drawn from the staff of Autonomics Division, were divided into four equal teams who worked with all three groups of

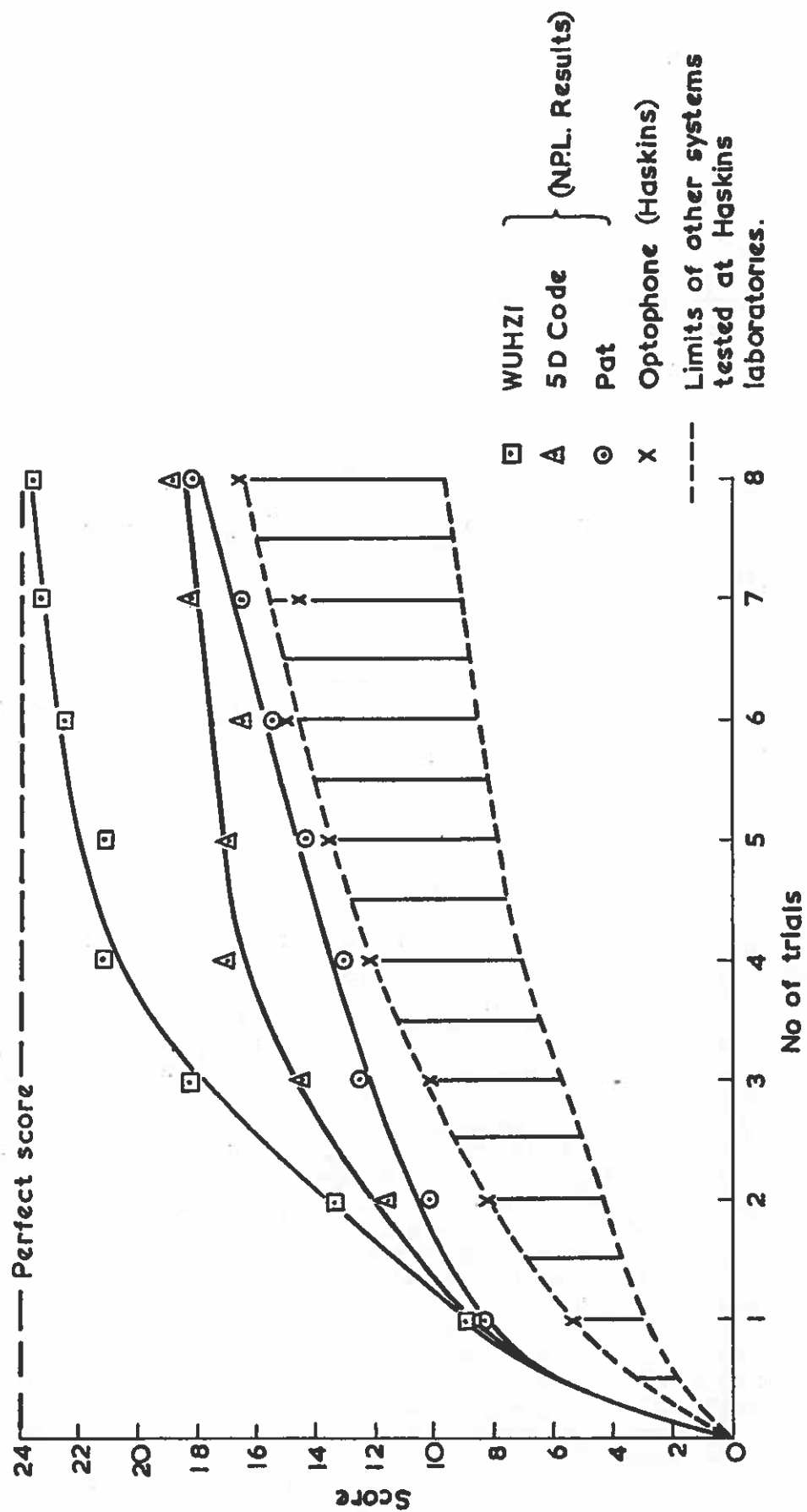


FIG.14.

word-sounds in different orders. The training technique followed the "Haskins Method" and the learning curves that were produced are shown in Fig. 14. The curve for the code Wuhzi coincided exactly with that traced by the Haskins Research Workers and it was therefore reasonable to assume that the conditions of their experiment had been accurately replicated. For this reason the results obtained at Haskins Laboratories with the optophone and a number of other systems have been plotted on the same graph for comparison purposes.

In Table II the results are shown of a Chi squared test made between the "observed" results obtained with the PAT and MDO displays and the expected performance, which for the present purpose has been taken to be that of the optophone. The MDO curve is found to be significant to rather better than the 1% level, whilst the difference between the PAT curve and the optophone could easily arise by chance.

TABLE 11

A Chi Squared Test on data from Fig. 14
No. of degrees of freedom = 7

Output	Chi Squared	Significance
MDO	20.19	P = 0.007
PAT	3.74	P = 0.81

6.9 There are two observations drawn from Fig. 14 that are worth noting. Firstly, the learning curve for the PAT output rises more slowly than the MDO curve and, secondly, extrapolation beyond the eighth series predicts that the curves will cross over and the PAT output will give an ultimately superior performance. A possible explanation of the crossover may lie in the fact that the sounds differ in their basic structure. In particular, the PAT sounds are smoothly varying, whereas the MDO sounds contain a number of transients; typically when the frequency switches from 480 cps to 2000 cps. Thus the MDO sounds have a marked rhythm which may alone carry sufficient information to specify eight alternatives. One can proceed to speculate that there are two mechanisms involved. The first controls the discrimination of stress and timing and functions quickly thus, the first features that are memorised are the transients. With further exposure to the sounds a second mechanism may come into play and enhance the discrimination process; for example, if the mechanisms of the feedback theory do exist and the sound can be mimicked, it might take time for a reference link to be established between the acoustic pattern and the listener's own verbal representation of the sound. Equally, as shown below, other mechanisms can be postulated that would achieve the same result.

A possible explanation of the slower learning performance with the PAT sounds might be that arising from the observations of House et al. (10). These workers found that for a series of sounds, ranging from unidimensional to multidimensional and finally to speech stimuli, the performance during learning improved as the number of dimensions increased but, as the machine-generated stimuli became more like speech, performance deteriorated. However, for natural speech signals, performance was at once superior to all other stimuli. The reason suggested by House et al. was that the synthesised and natural speech sounds are interpreted as linguistic events and are discriminated by an entirely

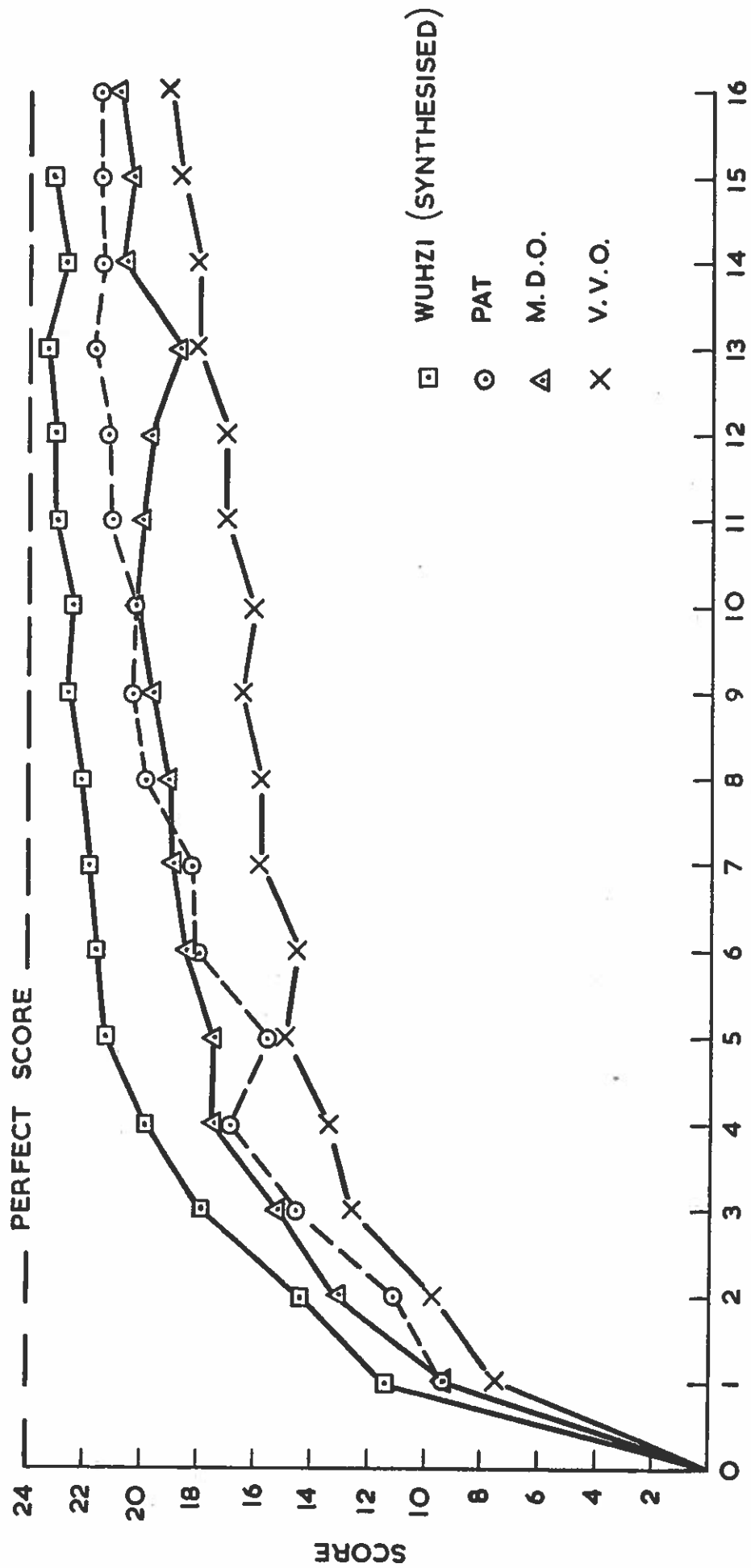


FIG. 15.

different mechanism to the multi-dimensional stimuli. All the speech sounds are therefore presented within a linguistic frame and are categorised accordingly. It is because the synthesised speech fails to fit this framework precisely that the performance is consequently poorer.

6.10 To investigate whether a crossing of the PAT and MDO curves would actually occur if the experiment were extended beyond the eighth series, a longer version of this experiment was performed on a second group of subjects. In addition, to obtain some indication of whether the phenomena reported by House et al. were active in depressing performance with synthesised speech sounds, the Wuhzi vocabulary was replaced by the same words synthesised on the PAT apparatus. The quality of articulation of these words was very much poorer than those produced by the human speaker in the previous experiment, and it provided an opportunity to compare the performance of synthesised and actual speech under conditions which by no means favoured the synthetic speech sounds. Included with the Wuhzi, PAT and MDO sounds was a fourth set of word sounds generated by a variant of the optophone in which the output intensities of the six tuned oscillators could be controlled by the same continuous functions that controlled the PAT and MDO output devices. This output is termed the variable volume optophone or VVO.

6.11 Twenty-four paid subjects, aged between 16 and 25 years, were engaged and given the task of learning all four sets of word sounds. The subjects were grouped in pairs and order effects reduced by the use of an orthogonal 4 x 4 latin square design. Two subjects resigned before the experiment was completed and the results of the remaining twenty-two were pooled to give the data plotted in Fig. 15. The performance of the synthesised Wuhzi, over the first five series, is identical with that found earlier with the naturally produced word sounds. Beyond the fifth series the curve does not continue to rise as steeply and the perfect score level is achieved with consistency by about 18 of the 22 subjects. The results of Chi squared tests of the differences in performance between the VVO sounds and the MDO, PAT and Wuhzi outputs in turn are shown in Table 12. These provide only a rough test of significance because the method of calculation does not eliminate subject variances.

TABLE 12

A Chi squared test on data from Fig. 14
No. of degrees of freedom = 15

Output	Chi Squared	Significance
MDO	7.76	P = 0.93
PAT	8.33	P = 0.91
WUHZI	24.45 (df = 14)	P = 0.03

6.12 A comparison of the results of the two experiments suggests that the differences between the performances are more reliable than the Chi squared test indicates. For example, the slower initial learning rate for the PAT sounds compared with the MDO output is repeated in the second experiment and the cross over predicted by the earlier data does indeed occur. It is also interesting to note that both the MDO and PAT outputs permit better performances than the VVO system, despite the fact that all three systems are controlled by the same waveforms and therefore receive the same input information. One other incidental

observation is that some of the information contained in the VVO is more clearly represented than in the PAT output. This is particularly true for the waveform which controls the third formant of the speech-like display. Variation of formant 3 makes a barely discriminable difference to the output even in a relative judgement situation. Thus, despite the loss of this source of information the sounds are still superior to the VVO output. The choice of the control parameter allocation for the PAT output used in these experiments was largely ad hoc (see Section 4.8) and it is clear that a significant improvement would result from a reallocation of the "Ribbon Height Vertical" control to some other more discriminable feature of the speech display.

It is possible to draw three conclusions from these results. Firstly, that the data are consistent with the view that the introduction of transient or consonant-like sounds can increase the rate of learning particularly in the initial stages. Secondly, in agreement with House et al., the multi-dimensional outputs can increase the rate of learning but, finally, there is no evidence that synthesised speech sounds lead to a measurably slower rate of learning than can be obtained by the same methods from natural speech.

6.13 Apart from the differences in learning time, the Haskins and N.P.L. training methods differ in the manner and speed of the feedback. With the Haskins method, feedback is received in a matter of seconds, but, in general, with the N.P.L. method the responses are corrected minutes later when a practice test of 20 sounds has been completed. Studies of the effects of delay of information upon learning have shown that quite dramatic changes in performance can occur and, to investigate whether the results of the previous experiment owed anything to the training methods, a further experiment using the N.P.L. procedure was mounted with the MDO, PAT and VVO outputs.

6.14 Twelve paid subjects of the 26-25 age group were used on this occasion and the usual precautions taken to avoid order effects. A total of $3\frac{1}{4}$ hours were spent on learning each group of sounds and the results of the final tests for each output were pooled to form confusion matrices. The results of an informational analysis is shown in Table 13.

TABLE 13

Output	Information Transmitted	Maximum possible Information Transfer
MDO	2.24 bits/symbol	2.99 bits/symbol
PAT	2.21 bits/symbol	2.99 bits/symbol
VVO	1.55 bits/symbol	2.99 bits/symbol

A statistical analysis of the significance of the differences between the three outputs is shown in Table 14.

TABLE 14

No. of degrees of freedom = 11
The pooled results of 12 subjects

Output	Mean of No. Correct Max. = 40	Variance	"t"	Significance
1. MDO	33.8	25.7	between 1 & 2 0.28	P = 0.8
2. PAT	34.2	31.02	between 2 & 3 4.36	P = 0.001
3. VVO	27.3	53.58	between 1 & 3 4.38	P = 0.001

From Table 14 the average number of correct responses with the PAT display is seen to be slightly better than that from the MDO output, but Table 13 indicates that the amount of information transmitted by the PAT system is nevertheless less than the MDO. The reason for this apparent disagreement arises from the fact that figures for both the variance and the number of correct results are employed in the calculation of information transmitted, and the variance of responses to the PAT output is higher than to the MDO output. The differences between the PAT and MDO outputs and the VVO sounds are highly significant and, although the PAT performance is still slightly better than the MDO, suggesting that the different training technique does not affect the performance, this difference is not significant. An analysis of the principal sources of variance is shown in Table 15. "Codes" refers to PAT, MOD and VVO outputs.

TABLE 15

Analysis of Variance

	Sum of Squares of Deviations	Degrees of Freedom	Mean Square Deviation
Codes	2160.1	2	1080.05
Order	1844.9	5	368.98
Tests	239.7	11	21.78
Codes x Order	830.8	10	83.08
Codes x Tests	706.8	22	42.13
Order x Tests	3051.0	55	55.47
Codes x Order x Tests	1371.0	110	12.46

The results of a Z test on the data of Table 15 are shown in Table 16.

TABLE 16

	Z	Degrees of Freedom		Probability of Variance arising from Random Sampling
		n ₁	n ₂	
Codes	2.23	2	110	P < 1%
Order	1.69	5	110	P < 1%
Tests	0.28	11	110	5% < P < 10%

The analysis shows, as indicated by the previous experiment, that the differences in performance with the codes is highly significant. Both order and subject variances are contained under the entry labelled order, but, because order effects have been made small by the latin square arrangement, subject variations represent the largest contribution. The Z test shows that these variations are significant at little better than the 1% level; also the variation in speed of presentation of the final tests had no significant influence on the performance.

6.15 In Section 2.9 an alternative to the feedback theory was presented which attempted to explain some features of the responses to vowel and consonant stimuli. To test this theory an experiment was designed which endeavoured to find out whether the same kind of discrimination behaviour could be demonstrated with non-speech sounds mapping similar trajectories in the frequency - time plane.

A frequency modulated oscillator (Appendix C) was used to generate two series of thirteen stimuli, each lasting for 400 mS. The first series was composed of continuous tones spaced at equal logarithmic intervals in the range from 300 cps to 1200 cps. The second series of sounds (the transients) traced a path in the frequency-time plane similar to that followed by the second formant of the /b/ /d/ and /g/ consonants shown in Fig. 1., but were sited centrally in the 300 cps - 1200 cps frequency band. Thus, the transient sounds consisted of a changing frequency, 100 mS in length, emanating from one of thirteen equally spaced points along the frequency band and terminating in a continuous tone at 750 cps which lasted for the remaining 300 mS.

6.16 Twelve volunteer subjects from the laboratory were recruited and given the task of, firstly, making relative discriminations of the sounds presented in an ABX mode and, secondly, of classifying the sounds into three groups called Low Medium and High.

An ABX presentation consisted of any two different stimuli A and B selected from the series of tones or transients followed by the sound X whose true identity was either A or B. Thus, the subject's task was to identify sound X with either A or B and write down the appropriate letter. Three sessions each consisting of ten series of 20 trials were administered to the subjects. Session one tested the discrimination of adjacent sounds from the range, also sounds separated by two stimuli all in random order. Session two presented sounds separated by two three and four stimuli and session three included sounds separated by four five and six stimuli. Average numbers of correct discriminations for each. Session for both Tone and Transient stimuli are shown in Table 17.

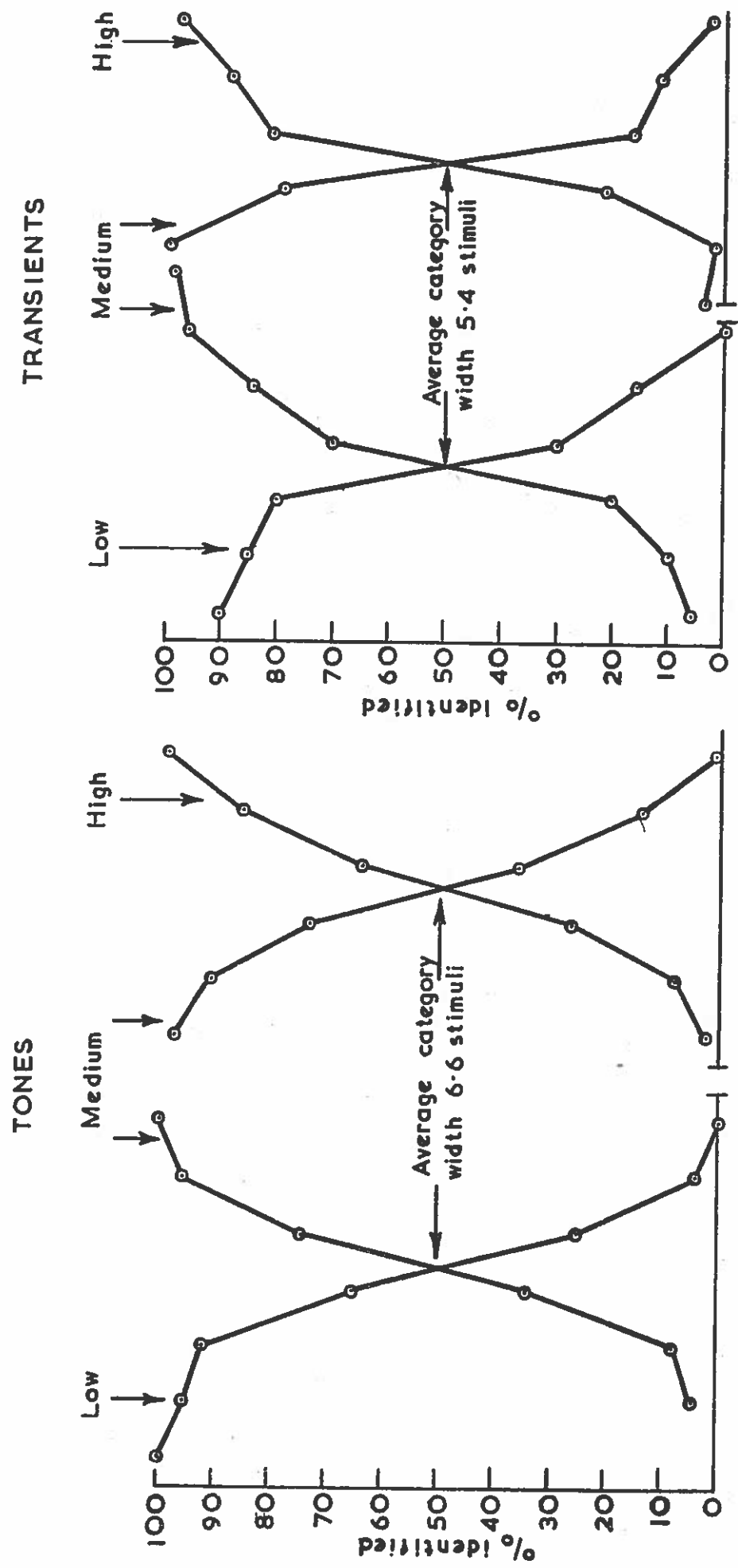


FIG.16.

TABLE 17

Stimuli	Session 1	Session 2	Session 3
Tones	97%	99%	100%
Transients	67%	82%	91%

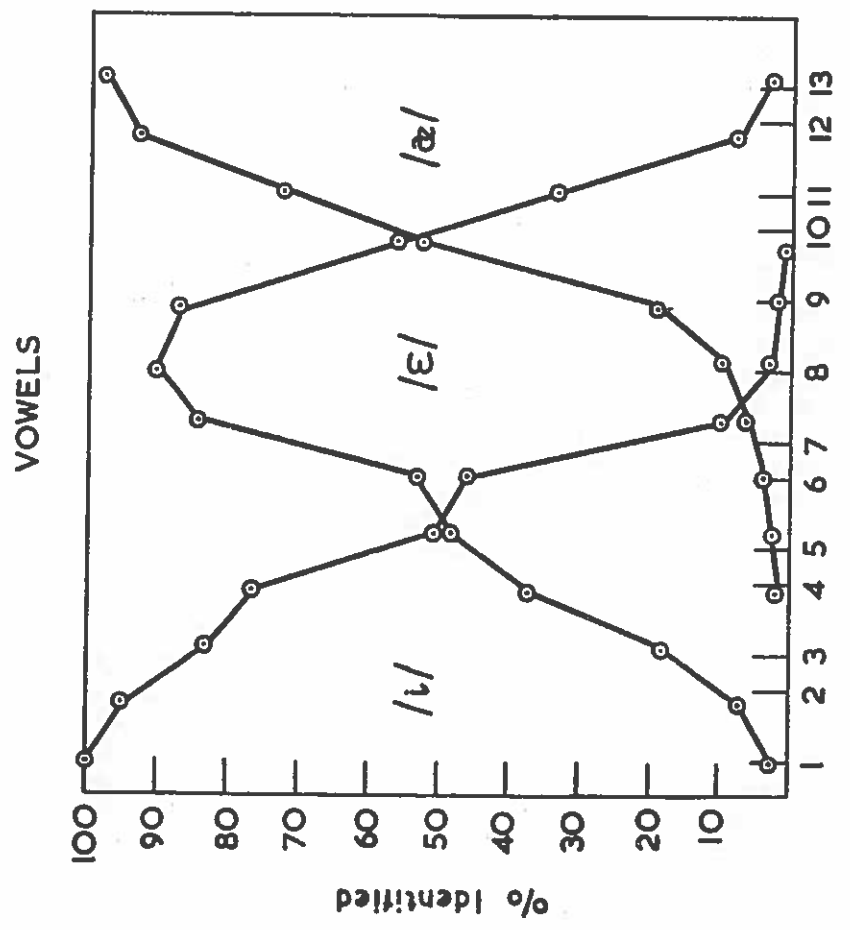
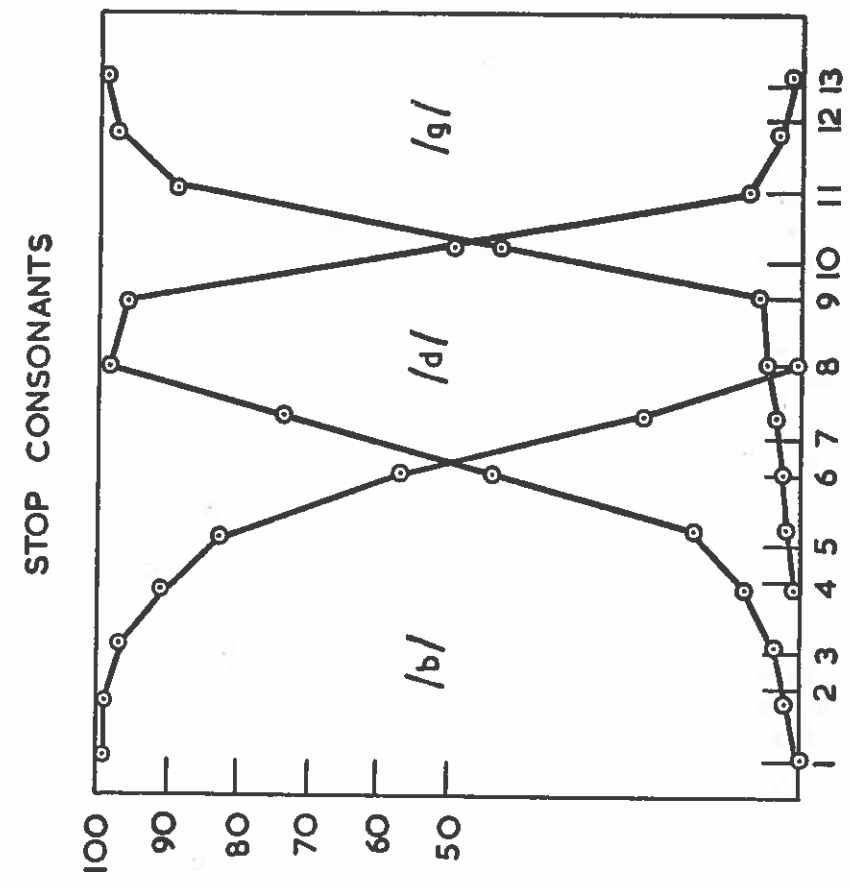
As the stimuli A and B are drawn from more widely separated points in the range so discrimination becomes more accurate. Table 17 also shows that relative discriminations can be carried out more easily with tonal than with transient stimuli.

In the classification experiment the tones were categorised according to three pitch levels whilst the transients were categorised according to whether the transitions originated from Low, Medium or High points in the frequency band. The subjects were instructed to call the first four sounds Low, the five sounds in the middle of the range Medium and the remaining four sounds High. The preliminary instruction was concluded by playing the entire range of sounds through once and following this by the three groups of sounds prefaced by the category names. The test recordings consisted of 10 groups of 20 sounds each separated by a three second interval. Having received their instructions the subjects listened to the two test recordings and wrote down their decisions on a response sheet.

On the basis of their skill at categorising both kinds of stimuli six subjects were selected and re-tested under better acoustic conditions in a sound-proof room. Twelve groups of twenty sounds were administered. The first two groups were provided for training purposes and lists of correct responses were supplied. The construction of both Tone and Transient test recordings was the same.

6.17 None of the subjects produced results which entirely contradicted the predictions made in Section 2.9 but in only two cases were the results unmistakably favourable. A composite graph shown in Fig. 16, in which results were not averaged directly, was constructed because the subjects did not place their boundaries at the same point along the stimulus range. Thus Fig. 16 shows the average gradient at the boundaries after alignment. The average width of the medium category has been drawn to approximately the correct proportion and shows that at the 50% level it was 5.4 stimuli for transients and for tones 6.6 stimuli compared with the 5 stimuli demanded in the initial instructions. The more accurate positioning of the transient category boundaries was reflected in all the subjects' results.

If the boundaries of Fig. 16 are inspected between the 20% and 80% levels the differences between the gradients indicate that the Transients were classified more precisely than the Tones. However, these differences are not as marked as those demonstrated for the vowels and stop consonants in Fig. 17. The reason for this may be in differences in the length of training or possibly it may arise from an observation made by nearly all the subjects. This was that the Transient stimuli were by far the more difficult to identify because it was very easy, during a moments inattention, to miss the initial transient. The Tones on the other hand allowed considerably more time for discrimination. Comparison of the conditions for Transient detection with those for stop consonants (described by Eimas (6)) shows that the listener to the speech sounds could be utilising the short voice bar, shown immediately preceding the second formant transition in



VOWEL DATA FROM FRY et al (1962) and STOP CONSONANT DATA FROM EIMAS (1963)

FIG. 17.

Fig. 17, as a cue to the time of onset of the transition. Thus his attention could be arrested at the critical time.

The results of this preliminary experiment do not allow a rejection of the feedback hypothesis. There are still a number of unanswered questions that merit further investigation but, on the basis of the present evidence, the balance of probability must now weigh more heavily against the feedback interpretation.

Chapter 7

7.0 Concluding Remarks

Taking the three experiments described in Sections 6.1, 6.3, 6.7, 6.10 and 3.13, together the overall conclusion is that the MDO and PAT outputs give performances which are significantly better than the optophone or VVO signals. Just how much this improvement represents in increased reading speed is an important question but difficult to answer without introducing some gross assumptions. One vital piece of information which is lacking is the speed of response to each of the displays. As the Z test shows, the subjects were not pressed for time in making their responses. Thus, any estimates of possible reading speeds made at this stage will be insubstantial. In this context it is therefore worth warning against the conclusion that an output whose performance lies midway between the optophone and natural speech will necessarily give a reading speed lying midway between 20 and 200 wpm. The reason for this may be obvious; it is that the sample of stimulus material used in the experiments has been extremely small. If the number of stimuli were raised from 8 to 80 or 800, the relative performance from the MDO and PAT systems might slump dramatically. There is no way of being certain of this except to perform the experiment. At this point the scale of the undertaking rapidly becomes too large to be conducted in a normal laboratory and takes on the status of a field trial with a likely increase in the amount of work and a reduction in the rate at which information is gained. The alternative is to allow the direction of research to be guided by intuition and the introspections of unbiased subjects.

7.1 Many subjects have pointed out, quite correctly, that the sounds produced by the PAT device represent only a small selection of the total repertoire found in human speech. In terms of some n-dimensional perceived signal space (PSS) the PAT sounds are not widely separated and occupy a relatively small volume of the space. The subjects usually refer to the fact that the output fails to include certain vowels and consonants. Despite further developments of the PAT system leading to the generation of a more multi-variate output it is possible that introspective observations may still yield responses stated in the same terms. This may arise because the rapid transitional consonant-like sounds that the device will produce will initially compare unfavourably with familiar English consonants. Subsequent training would be expected to make the consonant-like and vowel-like sounds equally discriminable.

7.2 However, adults have only a limited capacity for learning completely new discriminations. This capacity depends in a complicated way upon various incentives and the learning time required to achieve a given performance level. It may prove necessary to reduce the learning time by making the machine generate a larger number of identifiable phonemes. The task of learning this output would then become no more difficult than that of learning a natural language. If the production of identifiable phonemes should prove necessary this will impose quite severe constraints on the control functions and it will not be possible to generate these waveforms from print without a considerable increase in processing complexity. From this point there is but a relatively small step to the machine providing spoken English. However at the present time there are no firm indications that these problems will necessarily arise.

7.3 It is the writer's view that there is no evidence which can contradict the conclusion that the principal factor controlling the efficiency of an audible display is its dimensionality. It also appears likely that there is no intrinsic reason why, given the same amount of training, the same ultimate reading speed could not be achieved with both speech-like (mimicable) or non-speech-like multi-dimensional outputs, provided that they do not violate certain physical resolution limits of the ear. However, the results of these experiments indicate that, if the output is generated by a speech synthesising machine, we can exploit some of the discrimination skills already acquired in natural speech communication and achieve better performances. However, there is some evidence that the PAT can emit sounds which contradict certain speech habits

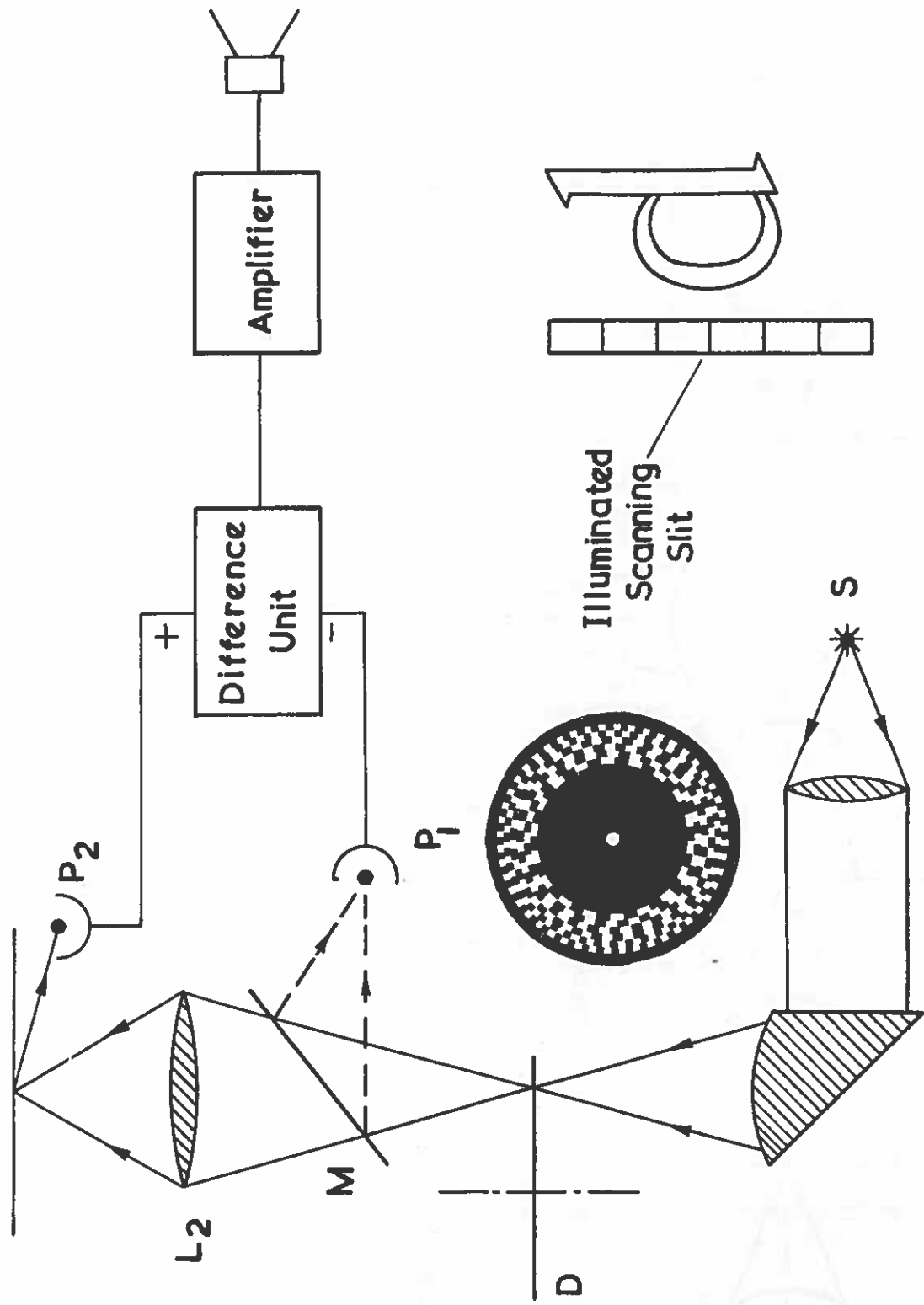
which have then to be unlearned. A good example of this is the case when a sharp pulse of noise, indicating an ascender, occurs in the middle of a long voiced sound. The impression given to the listener is that the sources of the voiced sound and noise pulse are quite separate and he is inclined to disregard the noise as interference. Thus the reader has to be prepared for combinations of sounds which do not occur in natural speech. It is to be hoped that effects of this sort will not prove to be insurmountable because it is highly probable that the avoidance of these situations by generating sounds which make even greater use of acquired speech discrimination skills, can only be attained at the cost of a more complicated processing logic. This will be necessary to limit the range of controls for the synthesizer to the form required to produce quasi or actual phonemes.

7.4 The scale of this research effort is modest and there are insufficient resources and time to continue to investigate deeply into the characteristics of a wider variety of multi-dimensional outputs. These experiments have shown the advantages that familiarity confers on speech-like outputs compared with non-speech displays and, despite the possible difficulties, the prospects of success seem to favour speech-like outputs. The present stage of development of the PAT system is rudimentary and there are a number of modifications which could be made which would lead to a more varied repertoire of sounds, and (probably) better performances in learning trials similar to those described here. With all reading machines, at some point the decision has to be made whether further development based upon introspective observations should be halted in favour of carrying out more realistic trials with larger vocabularies. Whilst it is too early yet to proceed to this stage with the PAT system, large scale evaluation of output along the lines adopted with the Battelle reader should not be delayed for too long, for it will provide the only convincing measure of feasibility.

References

1. Abma, J. S. Report on: The Auditory Encoding of Printed Characters. Battelle Memorial Institute, December 31st (1961)
2. Anthony, J. and Lawrence, W. A Resonance Analogue Speech Synthesiser. Proc. Fourth International Congress on Acoustics, Copenhagen, August 1962.
3. Beddoes, M. P., Belyea, E., and Gibson, W. C. A Reading Machine for the Blind. Nature 190 874 (1961)
4. Beurle, R. L. Electronic Aids for Blind People, London, St. Dunstan's (1952)
5. Cooper, F.S. and Zahl, P. A. Report On: Research on Reading Machines for the Blind. Feb. 1944-June 1947, Haskins Laboratories, New York City.
6. Eimas, P. D. The Relation between Identification and Discrimination along Speech and non-Speech Continua. Language and Speech 6 206 (1963)
7. Fry, D. B., Abramson, A. S., Eimas, P. D. and Liberman, A. M. The Identification and Discrimination of Synthetic Vowels. Language and Speech 5 171 (1962)
8. Garner, W. R. An Informational Analysis of Absolute Judgements of Loudness. J. Expt. Psychol. 46 373 (1953)
9. Hick, W. E. On the Rate of Gain of Information. Quart. J. Expt. Psychol. 4 11 (1952)
10. House, A. S., Stevens, K. N., Sandel, T. T. and Arnold, J. B. On the Learning of Speech-like Vocabularies. J. Verbal Learning and Verbal Behaviour 1 133 (1962)
11. Lawrence, W. The Synthesis of Speech from Signals that have a Low Information Rate. Communication Theory, Ed. Willis Jackson, pp. 460, Butterworths, London (1953)
12. Liberman, A. M., Harris, K. S., Hoffman, H. S. and Griffith, B. C. The Discrimination of Speech Sounds within and across Phoneme Boundaries. J. Expt. Psychol. 54 358 (1957)
13. McGill, W. J. Multivariate Information Transmission and its relation to Analysis of Variance. "Human factors" Operations Research Laboratories Report No. 32, Washington D.C.
14. Miller, G. A. The Magic Number Seven, plus or minus two: Some limits on our capacity for processing information. Psychol. Rev. 63 81 (1956)
15. Nye, P. W. Aural Recognition Time for Multi-dimensional Signals. Nature 196 1282 (1962)
16. Nye, P. W. Reading Aids for Blind People - A survey of Progress with the Technological and Human Problems. Med. Electron. Biol. Engng. 2 247 (1964)
17. Pollack, I. The Information of Elementary Auditory Displays. J. Acoust. Soc. Amer. 25 765 (1953)
18. Pollack, I. and Ficks, L. Information of Elementary Multi-dimensional Displays. J. Acoust. Soc. Amer. 26 155 (1954)

19. Studdert-Kennedy, M. and Liberman, A. M. Psychological Considerations in the design of Auditory Displays for Reading Machines. In Proc. Int. Congress on Technology and Blindness. L. L. Clark Vol 1. p. 289 A.F.B. New York 1963
20. Uyehara, G. U. A Stream Following Technique for use in Character Recognition. I.E.E. Convention Record, March (1963)
21. Yngve, V. H. Computer Programs for Translation. Scientific American 206 No. 6 P.68 June 1962



L1 FIG.18.

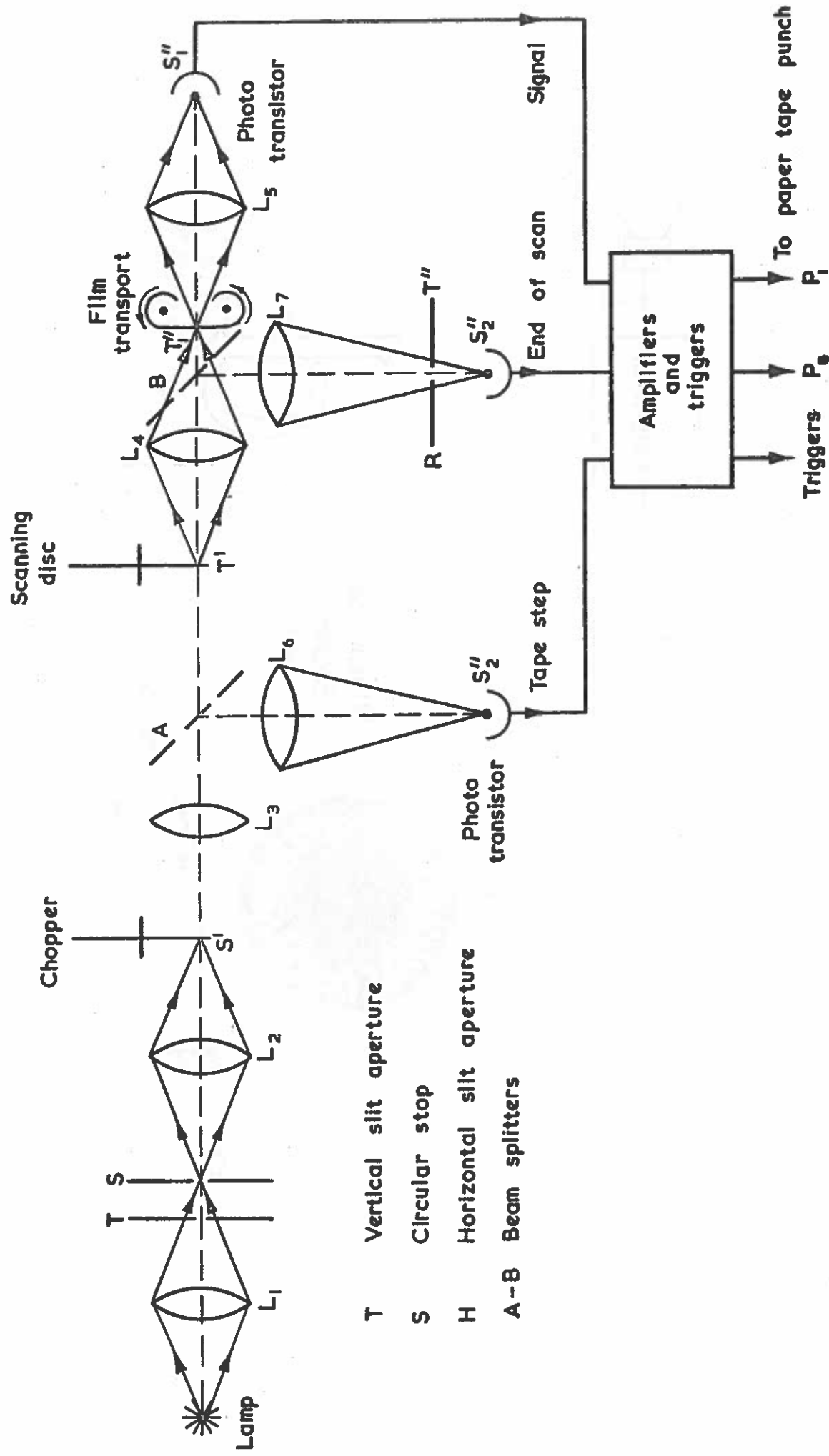


FIG.19.

APPENDIX A

The basic features of Fournier d'Albe's optophone are illustrated in Fig. 18. A rectangular slit image of the source is focussed with the long axis along the radius of a rotating sectored disc at D. Six annular rings on the disc modulate the transmitted light at different frequencies lying in the audible range. The lens L_2 focusses this light to form on the printed page an image of the slit at D. Reflected light from the page falls onto the photocell P_2 whilst a portion of the transmitted light from D falls onto a second photocell P_1 . Both signals are fed to a difference unit which is adjusted so that when white paper is placed in the reading position the signals balance and no output is heard. When a black region of a character falls under a portion of the illuminated slit the signals arriving at the difference unit become unbalanced and a tone, corresponding to the position of the black segment, is heard in the earphone.

A modern development of the optophone dispenses with the scanning disc and in its place uses a column of six or more photocells. Each cell is connected via an electronic switch to an oscillator tuned to one of a range of frequencies spaced at equal logarithmic intervals through the audio-spectrum. Following the convention of the Fournier d'Albe optophone the highest frequency is switched on by the top cell of the column and the lowest frequency by the bottom cell. In the case of Battelle reader output, which was simulated in these experiments, an oscillator is switched on when one third or more of a cell is covered by a character segment. The discrete switching action gives the output a staccato quality which contrasts sharply with the more smoothly varying display of the original optophone.

APPENDIX B

The scanning system which transforms print into digital form is shown in Fig. 19. A negative copy of the printed text is sited at the focus T_1^{11} between the beam splitter B and the lens L_2 and is transported horizontally. The lamp filament is focussed onto the circular aperture S, through a vertical slit T. S is focussed onto a chopper at S^1 having 36 blades and the slit T onto the scanning disc at T^1 containing sixteen radial slits. The scanning disc in its turn is focussed onto the film strip by L_4 and the transmitted light collected by the lens L_5 to fall on the photo transistor P_1 . This cell therefore provides a signal which depends upon the presence or absence of a character on the film strip. The chopper, scanning disc and film transport are all synchronised by a system of gears driven by a single electric motor. The two other photocells, Tape Step and End of Scan, are used to signal the sampling rate and the end of a scan. The horizontal slit R is placed so that only the last sample of each scan falls on the photocell P_8 . Finally the signals from these three cells are fed, via trigger circuits into a paper tape punch and the output subsequently processed by the feature analysis program. On average the width of a character is between five and six scans and each scan contains twelve one bit samples giving a matrix of between 60 and 70 bits per character.

APPENDIX C

With the exception of the Parametric Artificial Talking Apparatus which was designed by Lawrence (11) all the circuitry used in the stimulus generator was built in the laboratory.

The digital-analogue converter and de-multiplexing unit is shown in Figs 20, 21 and 22. Figure 20 shows a schematic diagram of the connections between the D-A converter, the six clamp amplifiers and the eight channel selector box. The output from the D-A converter is fed to all the clamp amplifier inputs. Signals

from the data selector (set for six channels) set each clamp amplifier in turn to the value appearing at its input. The counter of the data selector is maintained in phase by means of a resetting pulse derived from the eighth hole (H) of the paper tape. Figure 21 shows the D-A converter and a clamp amplifier. Figure 22 shows a semi-schematic diagram of the eight channel data selector unit.

Figures 23 and 24 show the circuits of the counter and tape sequencing unit. This device is used in conjunction with two tape recorders. The first machine carries an endless loop of tape on which the series of stimulus sounds has been recorded on track one and a series of short pulses at 1000 ops, preceding the sounds, recorded on track two. The signal pulse from track two triggers the circuit shown in Figure 23 and the output is fed into the counting circuit shown in Figure 24. The sequencing unit compares the number registered on the counter with one of a sequence of numbers punched on paper tape and when the numbers coincide the device switches on the second recorder which then records the stimulus sound. The device therefore assembles sequences of sounds drawn from the loop of tape in any order specified on the input paper tape.

Figures 25, 26 and 27 are circuit diagrams of some other devices mentioned in this report. Mr. D. L. A. Barber, Mr. J. R. Parks and Mr. E. P. H. Woodroff have been responsible for the design of much of the equipment described here.

APPENDIX D

Confusion Matrices

Section 6.1 Conventional Optophone

Stimuli Presented

	1	2	3	4	5	6	7	8
1	284	10	3	11	4	2	2	3
2	19	259	2	35	0	1	1	4
3	1	1	275	1	12	15	2	0
Responses 4	29	37	2	185	4	4	0	2
5	6	2	4	6	168	35	4	31
6	4	2	9	1	24	245	2	3
7	5	0	11	1	4	8	229	4
8	2	0	5	1	41	13	11	188

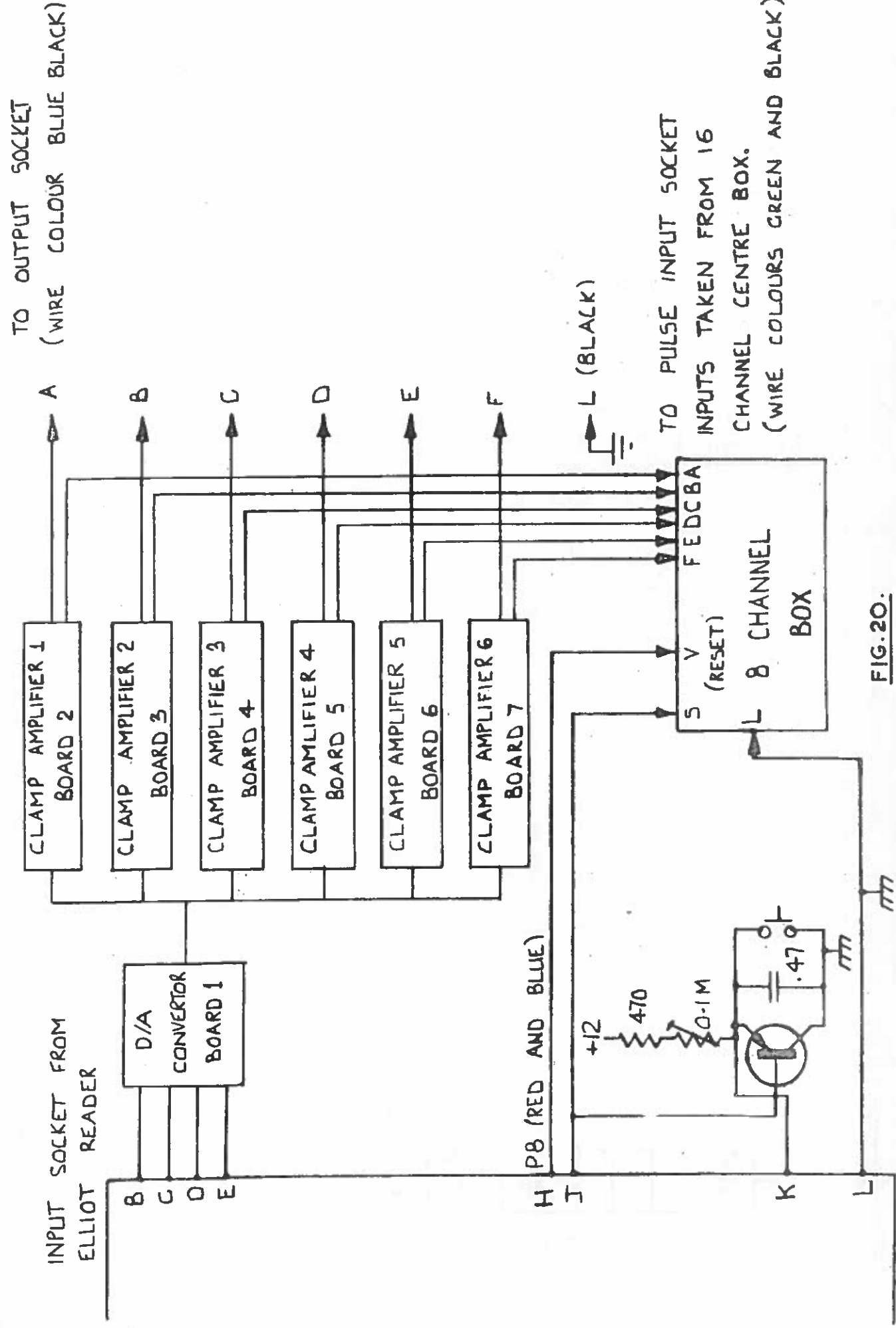


FIG.20.

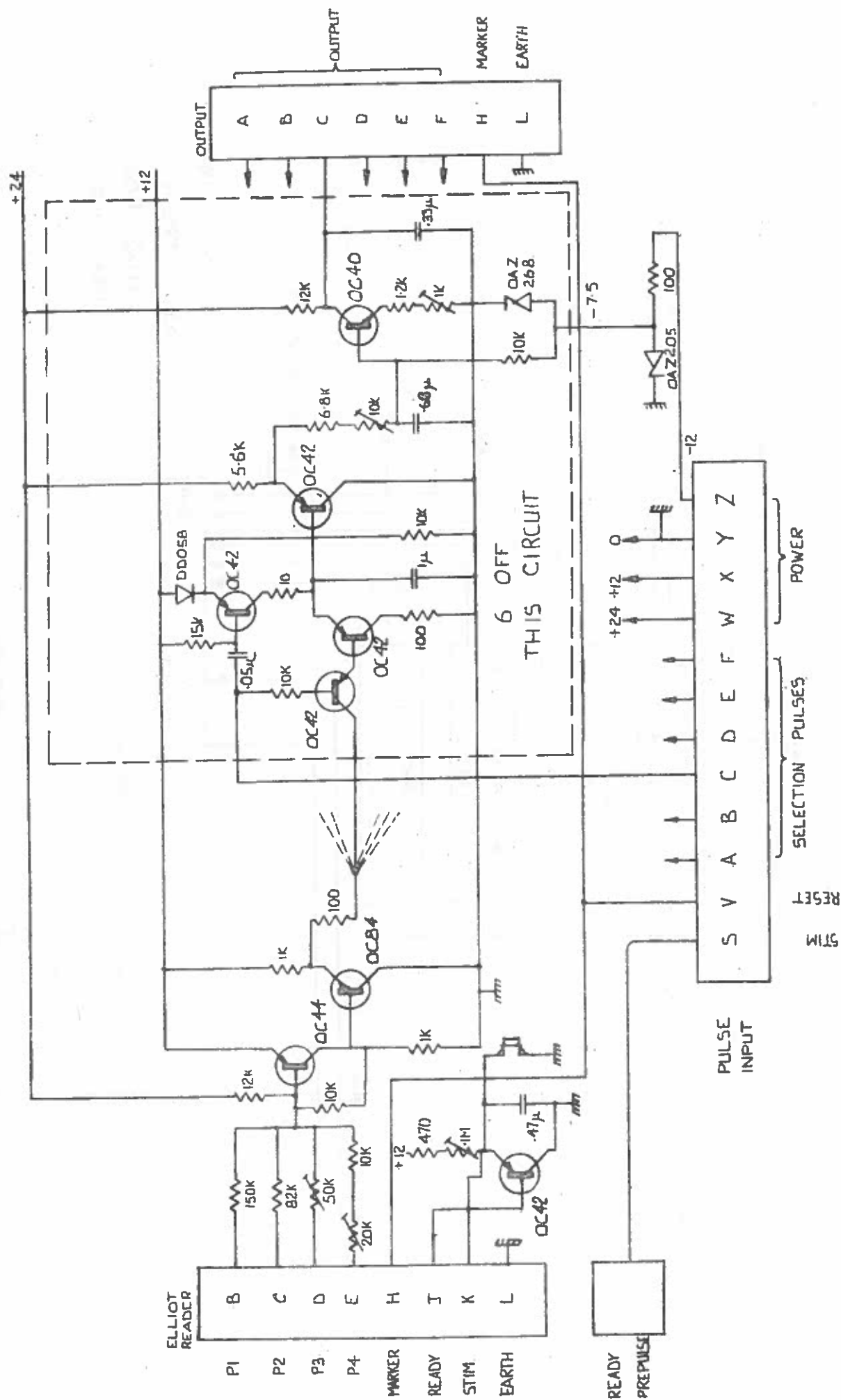


FIG. 21.

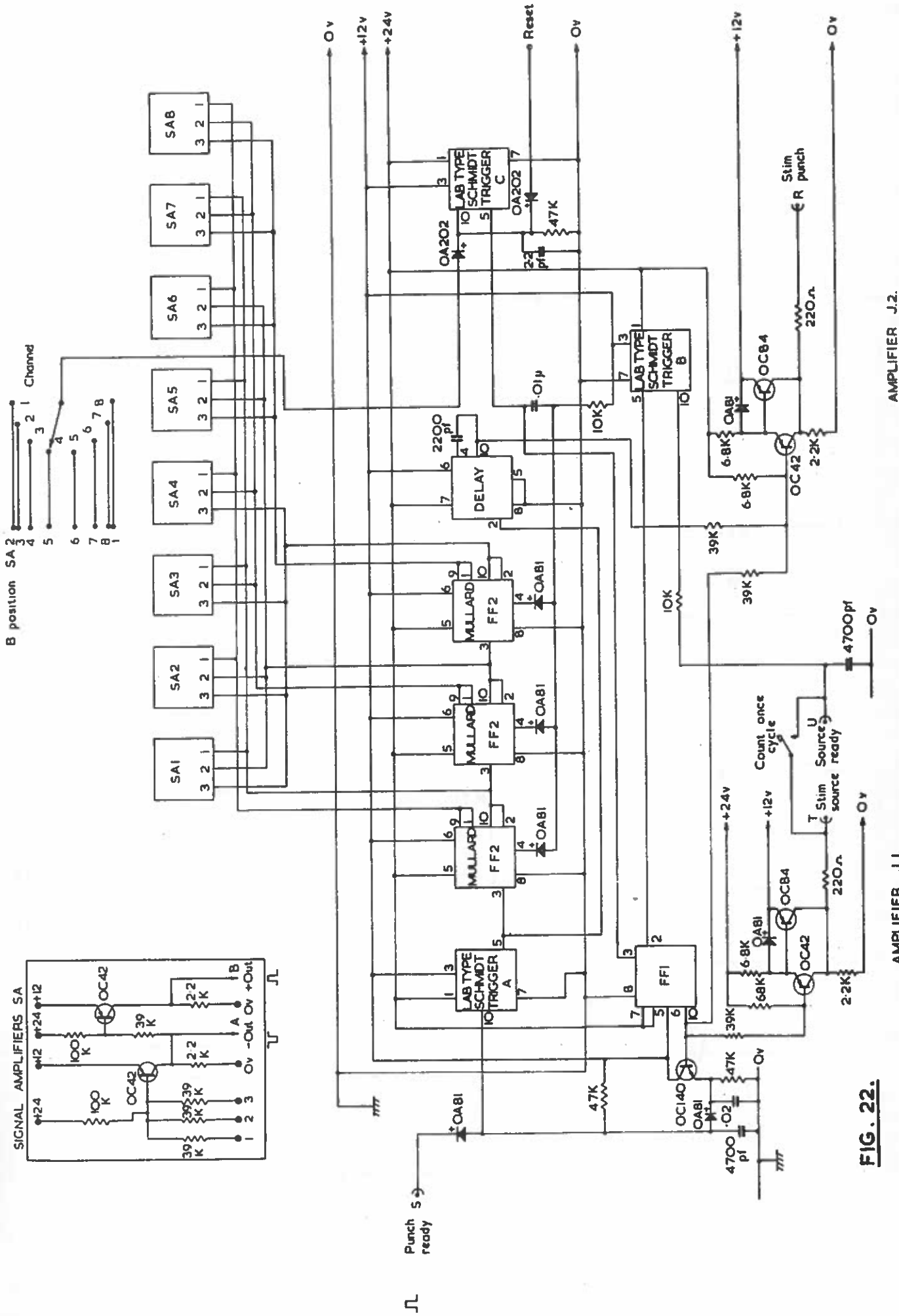
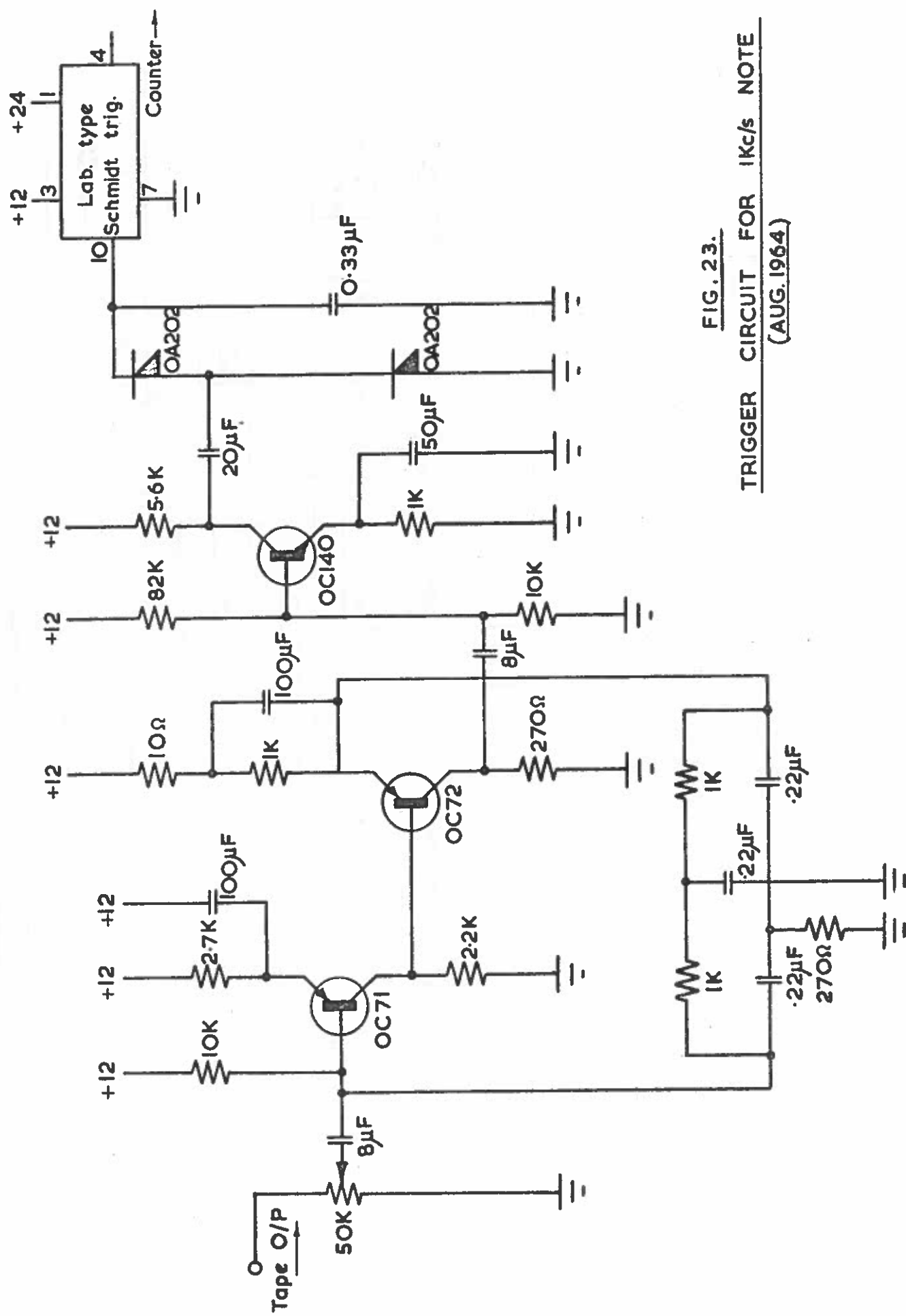


FIG. 22.

AMPLIFIER J 1

AMPLIFIER J 2.



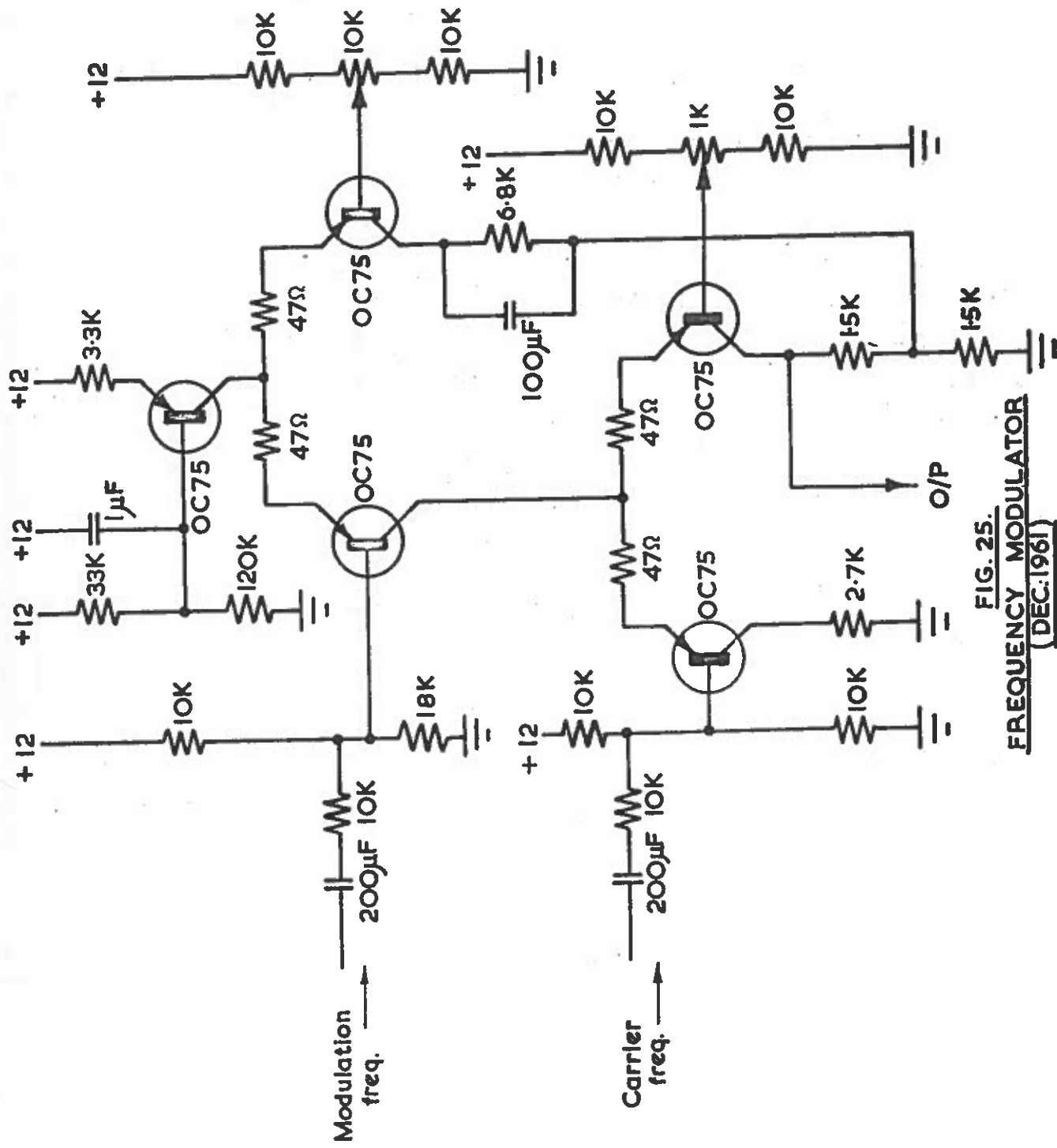
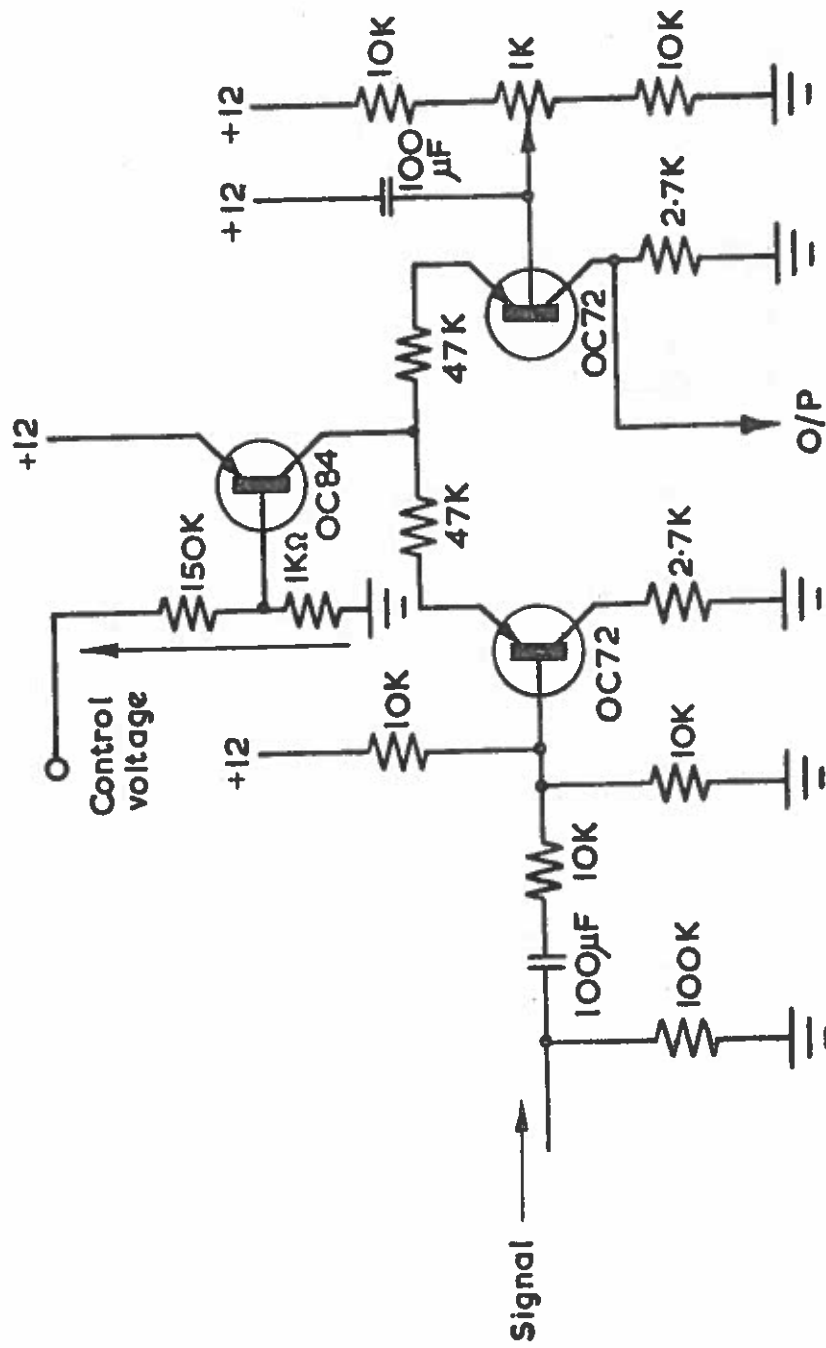


FIG. 25.
FREQUENCY MODULATOR
(DEC.1961)



VARIABLE VOLUME CONTROL
LOG. AMP. (CONTROLLED BY DCO-20V)
(NOV. 1963)

FIG. 27.

Compressed Optophone

Stimuli Presented

	A	B	C	D	E	F	G	H
A	388	12	2	2	0	0	0	4
B	16	276	0	1	3	3	3	2
C	0	0	315	0	1	5	0	1
Responses D	1	1	0	245	0	0	0	0
E	0	0	1	0	223	17	3	6
F	0	0	8	0	21	293	3	5
G	1	9	2	1	6	10	261	4
H	3	1	11	0	10	7	3	237

Section 6.3 Group 1 (Similar Sounds)

Stimuli Presented

	1	2	3	4	5	6	7	8
1	401	14	5	2	11	0	0	0
2	5	360	4	4	1	0	0	0
3	9	6	379	13	4	0	0	0
Responses 4	7	1	6	276	3	0	0	0
5	1	1	0	4	307	4	2	0
6	0	0	0	0	0	382	1	13
7	2	0	0	0	7	5	317	2
8	0	0	0	0	1	11	1	282

Group 2 (Dissimilar Sounds)

Stimuli Presented

	A	B	C	D	E	F	G	H
A	327	0	1	0	60	0	0	0
B	0	344	0	11	0	1	0	5
C	1	0	401	0	5	0	3	0
Responses D	1	24	0	280	0	2	0	2
E	94	0	5	2	246	0	1	0
F	0	0	0	0	0	398	0	1
G	9	1	0	2	10	2	330	0
H	0	1	0	1	0	3	0	287

Group 3 (Optophone Sounds)

Stimuli Presented

	S	T	U	V	W	X	Y	Z
S	396	2	2	0	3	1	1	5
T	3	374	1	5	2	2	4	1
U	12	0	316	3	0	3	3	28
Responses V	1	1	1	264	26	46	36	3
W	0	2	3	21	204	101	65	11
X	3	1	0	15	27	174	66	2
Y	1	2	1	15	51	75	99	3
Z	11	0	70	2	13	1	2	242

Section 6.13 PAT Output

Stimuli Presented

	A	B	C	D	E	F	G	H
A	392	6	16	4	15	1	0	0
B	4	313	5	10	11	5	4	2
C	14	1	346	3	16	1	1	0
Responses D	1	2	2	216	17	1	1	3
E	8	24	31	48	244	3	8	0
F	2	3	2	5	6	371	10	7
G	2	3	0	3	0	8	284	3
H	0	1	1	0	0	13	18	294

MDO Output

	I	J	K	L	M	N	O	P
I	315	57	1	9	4	1	0	0
J	44	286	0	5	1	15	3	0
K	1	0	361	7	3	0	0	0
Responses L	26	2	18	260	27	1	0	0
M	12	3	4	5	295	1	0	0
N	3	32	0	2	1	358	27	0
O	4	0	0	0	0	27	270	4
P	11	1	0	0	0	0	22	292