

SOME INVESTIGATIONS ON REDUNDANCY AND POSSIBLE BANDWIDTH COMPRESSION IN TELEVISION TRANSMISSION

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN
DOCTOR IN DE TECHNISCHE WETEN-
SCHAP AAN DE TECHNISCHE HOGE-
SCHOOL TE DELFT, OP GEZAG VAN DE
RECTOR MAGNIFICUS, DR. R. KRONIG,
HOOGLEERAAR IN DE AFDELING DER
TECHNISCHE NATUURKUNDE, VOOR EEN
COMMISSIE UIT DE SENAAT TE VERDE-
DIGEN OP WOENSDAG 23 SEPTEMBER 1959
DES NAMIDDAGS TE 2 UUR

DOOR

KEES TEER

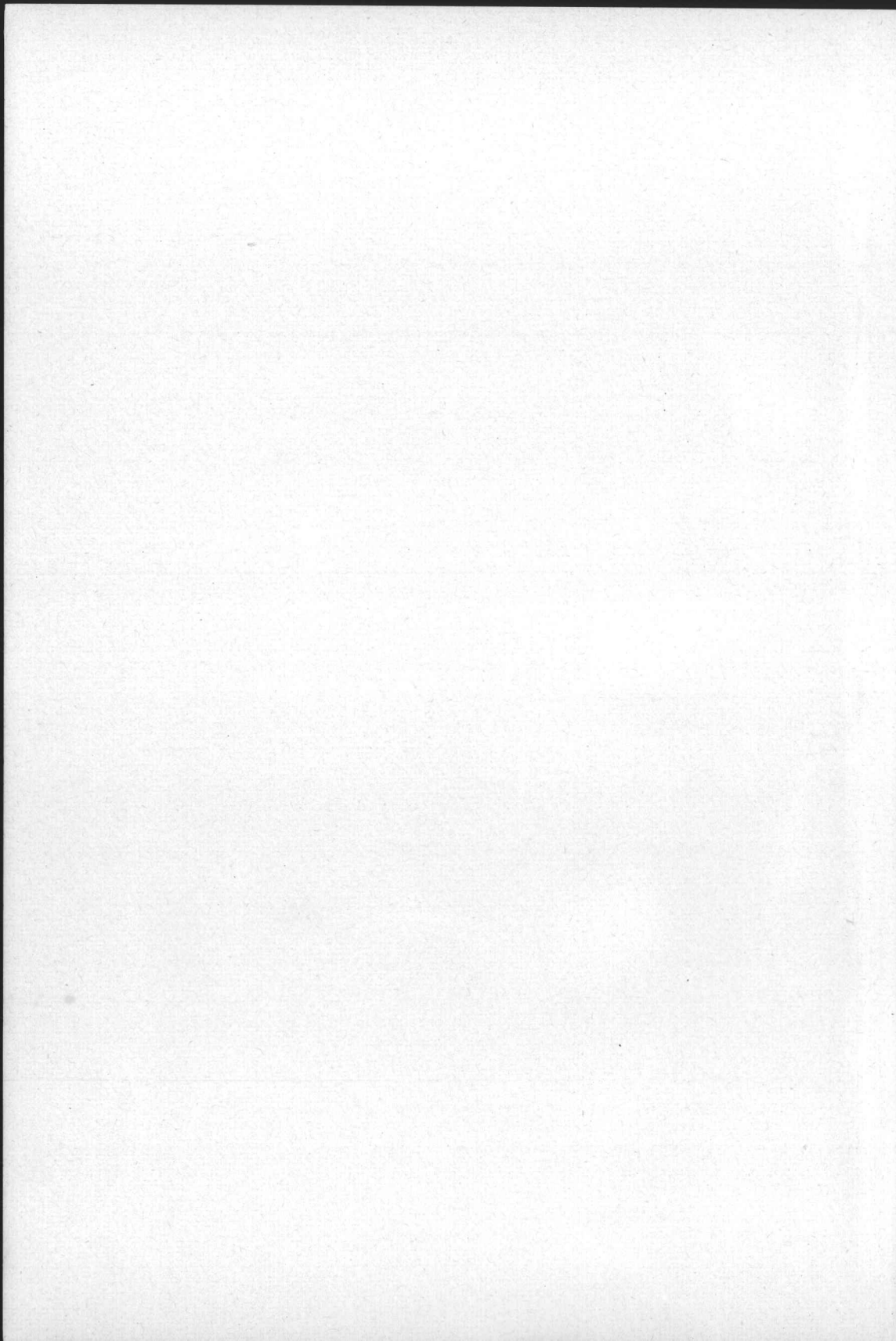
ELECTROTECHNISCH INGENIEUR
GEBOREN TE HAARLEM



H20
1012 ~~B160~~

DIT PROEFSCHRIFT IS GOEDGEKEURD
DOOR DE PROMOTOR
PROF. IR M. P. BREEDVELD

aan mijn ouders
aan mijn vrouw



CONTENTS

I. GENERAL INTRODUCTION

I.1. Outline of the present investigation	1
I.2. The relation between bandwidth and picture quality in a conventional television system	3
References	6

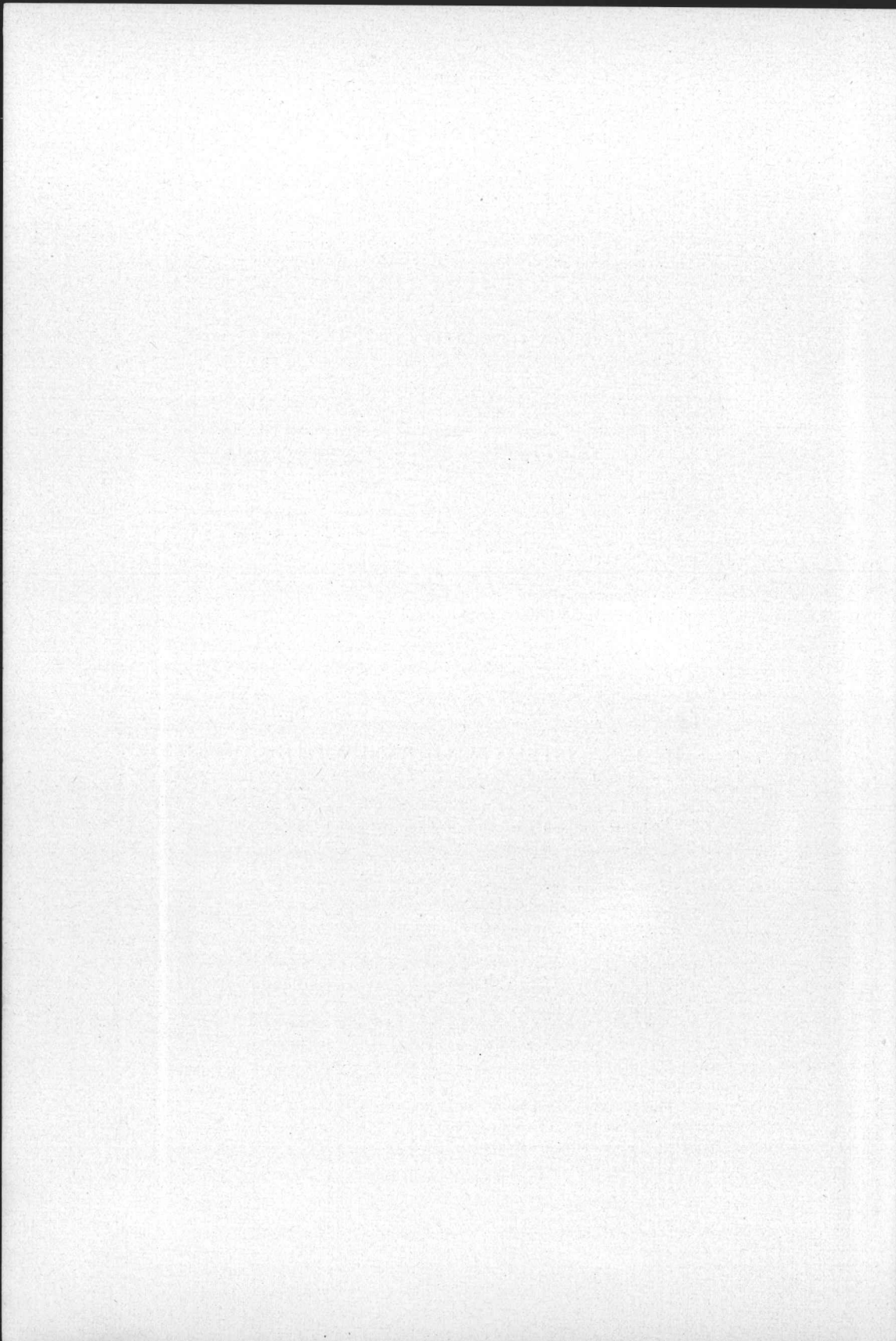
II. POSSIBILITIES FOR BANDWIDTH COMPRESSION

II.1. Introduction	7
II.2. Statistical redundancy	8
II.2.1. Correlation between adjacent picture elements	8
II.2.1.1. Measurements by Schreiber and Kretzmer	11
II.2.1.2. Measurement of the number of "essential picture elements"	14
II.2.1.3. Discussion	20
II.2.1.4. Systems proposed by Cherry and Gouriet	22
II.2.1.5. The effect of noise in the transmission of position data	23
II.2.2. Correlation between successive frames.	25
II.3. Physiological limitations	27
II.3.1. The resolving power of the eye	27
II.3.2. Persistence of vision	32
II.3.2.1. Discontinuity in motion	32
II.3.2.2. A decrease of the information per field	38
II.3.3. The differential sensitivity of the eye	42
II.3.4. Perception of colour	45
II.4. Psychological limitations	48
References	49

III. TRANSMISSION SYSTEMS WITH BANDWIDTH COMPRESSION

III.1. Introduction	51
III.2. A decrease in the number of fields per second	51
III.2.1. Theoretical analysis of the vidicon used as a memory device	55
III.2.2. Experiments on duplex transmission	62
III.3. More than two fields per frame	68
III.3.1. Dot-interlace	68
III.3.1.1. Realization of the transmission characteristic	74
III.3.1.2. The introduction of a third signal process	78
III.3.1.3. Picture reproduction	80
III.3.2. Subcarriers	86
III.3.2.1. Comparison of subcarrier system and dot-interlace system	88
III.3.3. Signal separation at the receiver	90
III.3.3.1. Non-synchronous detection of a subcarrier	90
III.3.3.2. Non-synchronous detection in a dot-interlace system	95
III.3.4. Conclusion	97
III.3.5. Experiments	98
III.4. The transmission problem in colour television	101
III.4.1. The N.T.S.C. system	103
III.4.2. The two-subcarrier system	104
III.4.3. Other transmission systems	113
References	115

SUMMARY



SOME INVESTIGATIONS ON REDUNDANCY AND POSSIBLE BANDWIDTH COMPRESSION IN TELEVISION TRANSMISSION

I. GENERAL INTRODUCTION

I.1. Outline of the present investigation

In image transmission as takes place in television the bandwidth of the signal carrying the information is of the order of 3.5-10.5 Mc/s, depending on the number of lines. Compared with other forms of communication, this can be said to be very large. When it is borne in mind that one television signal occupies as broad a band as about 500 audio signals the question arises why this large difference between visual and audio transmission should exist. One tends often to suppose that either a large difference exists with respect to the capacity of the visual and auditive perception channel or that for some reason a large amount of redundancy is present in the television transmission process. In our opinion one must be very cautious about making such statements, in order to avoid over-simplification. A complete treatment of the problem in terms of information rate, channel capacity and redundancy as defined in information theory is not possible because ultimately one has to do with rational and emotional impressions caused by the transmission and in most cases these impressions cannot be described quantitatively. Only where the message contains purely factual data, and where conceptions such as "picture quality" or "sound quality" are of minor importance, can the information apprehended by the observer be more or less exactly measured and a comparison made between the various forms of communication. In that case the conclusion can be drawn that only a slight difference exists between the information rate for aural and visual communication ^{1,2}).

For most applications of television the situation is not so simple either, because the factual data are of minor importance or because these data are hidden in a complex picture content from which they cannot easily be separated without human intervention. In one particular respect it is possible to point to a clear difference between aural and visual transmission. This is with respect to the directivity of perception in space. Unlike the situation for hearing, in the viewing of an image, perception is concentrated on a special part of the whole picture area. In viewing, more than in hearing, a choice is continuously made, from an extensive total image content. The possibility of making such a choice being inherent in visual perception, it is inevitable in image transmission that far more should be transmitted than one observer can perceive.

So where on the one hand it is a doubtful thing to decide for an exceptional large redundancy in the television signal based on straightforward considerations of information theory — we shall return to this subject in section II.4 — on the other hand it is not right to assume a priori that it is impossible anyhow to achieve considerable bandwidth compression by suitable modifications of the transmission process. The large frequency band occupied in the ether, the limited transmitter range, the very high requirements on links, all these make the wide band such a telling disadvantage, that it is well worth while to examine how far such modifications are possible.

Obviously it is not only the saving in bandwidth which determines the value of a new system. Apart from the required signal-to-noise ratio and the influence of the system on picture quality, very important aspects are the complexity of system equipment at either end of the transmission channel, the required transmission channel characteristics, the “compatibility”, and the susceptibility to interference.

In the following pages the problem of bandwidth compression in television will be studied. A wide interpretation is given to the term “bandwidth compression”, and according it covers also multiplex transmission within a relatively narrow bandwidth and bandwidth compression involving some deterioration of picture quality. In a single case there is no saving of bandwidth, but the signal-to-noise ratio is improved, so in this case it would be better to speak of a reduction of channel capacity.

The first part of the study is devoted to an analysis of those aspects of image content and of visual perception which may allow a certain reduction of transmitted data. Most of the considerations have an experimental basis. Sometimes reference will be made to redundancy and bits/second, notwithstanding what was said above. However, this kind of description of the impression is always followed by discussions of picture quality, disturbing effect, acceptability and, in general of concepts which emphasize our objections, as set out above, to using the terminology of pure information theory alone.

In the second part of the study, consideration is given to the practical design of transmission systems with reduced bandwidth along the lines laid down in the first part. It deals with the principles, the building up, the characteristics and the results of experimental systems. Colour television transmission is given special attention because so far the most successful attempts to narrow a television channel have taken place in the field of colour television.

For this study no claim is made for completeness. Experiments in perception are often of an empirical nature, allowing a quick decision to be made about the usefulness of a principle. Investigation of a system involves signal operations, pick-up and display processes which are often entirely new. We have not been able to go into all the inherent technical problems at length. Possibly a lack of quantitative results and conclusions may be noticed. However

this is also due to the difficulty, already referred to, of describing perception in terms of numerical values. An "acceptable picture" being a somewhat vague aim, and "picture quality" being impossible to measure, the lack of figures is inevitable.

Before embarking on the first part of our study, we shall do well first to bring out the connection between bandwidth and picture quality in conventional television systems. This will be done in the next section.

I.2. The relation between bandwidth and picture quality in a conventional television system

The image present at the transmitter side can be characterized by the function $B(x, y, t)$ which represents the dependence of the brightness on place and time. The task performed in television can be described as the transmission and visual display of the information in $B(x, y, t)$. It is not possible to transmit $B(x, y, t)$ with absolute exactitude. The practical difficulties increase rapidly with the exactitude required. However it is not necessary to strive for an unlimited accuracy because the observer at the receiver side will be satisfied if his impression corresponds to the original image to a sufficiently large extent *). It is permissible for the transmission process to exhibit special limitations corresponding to natural limitations to perception. In existing television systems the limits set to the accuracy and the reaction time of the human eye are involved, and apart from an accidental disturbance (noise etc.) a certain systematic inaccuracy in the reproduced image $B_R(x, y, t)$ is accepted which is matched to the characteristics of the eye. This makes it possible to limit the bandwidth.

The inaccuracy in question is of two kinds. In the first place there is inaccuracy in that the value of B_R for x_1, y_1 and t_1 does not merely depend on $B(x_1, y_1, t_1)$ but depends also on values of B in the immediate neighbourhood of the point x_1, y_1, t_1 . A second kind of inaccuracy exists that B_R only corresponds with B for a finite number of equidistant values of the independent variable; in between these points B_R has a more or less arbitrary form.

If x is identified with the horizontal direction in the television picture and y is identified with the vertical direction, it will be clear that in the normal television system the first kind of inaccuracy occurs in the x direction and the second kind occurs in the y and t coordinate. In the direction of the lines the shape of B is given continuously and the limited pass-band will cause a certain "smearing". In the vertical direction and in time, transmission is discontinuous because it is established in the form of a finite number of lines and a finite number of frames per second. However, the integration effect also occurs in

*) In information theory the conception "reproduction quality" was introduced by Shannon to serve as a quantitative description of the correspondence between the received message and the original message ³⁾.

the y and t direction. In the y direction it is caused by the finite width of the spot in the camera and picture tubes. In time it is caused by the storage time of the sensitive layer in the camera tube and the decay time of the phosphor in the picture tube.

The inaccuracy characterized by an integration effect in a region around the point concerned can be described by a convolution integral. When it depends only on one coördinate, here denoted as ζ , this gives:

$$B_R(\zeta_1) = \int_{\zeta_1 - \Delta''\zeta}^{\zeta_1 + \Delta'\zeta} B(\zeta) w(\zeta_1 - \zeta) d\zeta. \quad (1)$$

Hence integration takes place over the interval $\Delta'\zeta + \Delta''\zeta = \Delta\zeta$ around ζ_1 . Here $w(\zeta)$ is a kind of weighting function which determines the influence of surrounding points on the ultimate result, evidently being zero outside the interval. A corresponding two dimensional integral can be given for an inaccuracy depending on two coordinates. This description of inaccuracy applies for the influence of the electron-beam spot in camera tube and picture tube. However, generally it suffices to consider the accuracy in x , y and t separated and in that case the form (1) can be used.

Since scanning takes place horizontally the transmitted signal varies in time in accordance with $B(x)$, and for the greater part $B_R(x)$ is determined by the frequency characteristic of the transmission channel. So $w(x)$ corresponds to the unit-pulse response $w(t)$ of the transmission channel. As is well known the following relation exists between the unit-pulse response $w(t)$ and the transmission characteristic $\bar{W}(\omega)$

$$w(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \bar{W}(\omega) e^{j\omega t} d\omega. \quad (2)$$

The received signal $u_R(t)$ derived from the original video signal at the transmitter $u(t)$ is for $t = t_1$

$$u_R(t_1) = \int_{-\infty}^{\infty} u(t) w(t_1 - t) dt. \quad (3)$$

According to (3) $u_R(t_1)$ is determined by all values of $u(t)$ from $-\infty$ up to $+\infty$. Of course with a time variable it is impossible to get any influence of $u(t)$ on $u_R(t_1)$ for $t > t_1$. Hence only such functions $\bar{W}(\omega)$ are physically possible as correspond to a unit-pulse response $w(t)$ which is zero for $t < 0$. In fact $w(t)$ only contributes significantly to the integral in a limited region Δt , so that the integration interval can be limited to Δt .

At a certain instant t_v , $w(t)$ will show a maximum value. It is therefore the value of $u(t_1 - t_v)$ that does most to determine $u_R(t_1)$. Of course a comparison

has to be made between $u_R(t + t_v)$ and $u(t)$ — rather than between $u_R(t)$ and $u(t)$ — because this will give a closer correspondance; the constant time delay t_v can be accepted.

When the limitation in the transmission channel is of such a kind that $B(\zeta)$ is transmitted only for special values of ζ at equal intervals of ζ_p , then $B_R(\zeta)$ can be described as a product of $B(\zeta)$ and a sampling function $h(\zeta)$, where $h(\zeta)$ is unity at the sampling points and zero for other values of ζ . Hence

$$B_R(\zeta) = h(\zeta) B(\zeta). \quad (4)$$

As already stated, in vertical direction as well as in time a sampling process and a transformation like (1) takes place. The latter operation happens twice, before and after the sampling. The ultimate result B_R is therefore determined by a number of functions and values like $w(\zeta)$ and ζ_p , where ζ is x , y or t respectively. The form and magnitude of these will be such that they are subject to certain limits corresponding to the properties of visual perception. These tolerances cannot be formulated very precisely but in a first approximation they can be given as maximum values of y_p , t_p and of the regions of w_x , w_y and w_t .

It should be noted here that these values are also related to the movements of the observer's eye. In section II.3.1 we will go into the matter further. Here we would merely point out that when the eye follows some moving object in the picture, the integration effect of the camera tube may cause some visible unsharpness in motion though the integration time is well within the limits set by persistence of vision. If the eye does not move with the object this unsharpness is still present, of course, but then the impression is the same as that of watching the original picture.

There is no exact relationship between the above tolerances and the visual acuity and persistence of vision, as will be clear from the subjective nature of perception. This is illustrated clearly from the variety of television standards employed in different countries. Though television broadcasting has long been an everyday reality, the discussions about picture quality and television standards have not yet come to an end ^{4,5}). The determination of an exact relationship between visual acuity and the number of picture elements is not the only difficulty: a further one arises from the difference between reproduction in the horizontal direction (continuously) and the vertical direction (discontinuously with line-interlace). Based on equal horizontal and vertical separation of picture elements the required bandwidth is

$$f = \frac{1}{2} \frac{4}{3} \frac{1}{T} n^2 \frac{a_i}{a_l} \quad (5)$$

where

$4/3$ is the aspect ratio of the image,

T is the scanning period for a frame *),
 n is the number of lines per frame,
 $a_1 n$ is the number of lines carrying picture information,
 a_1 that part of a line carrying picture information.

When the first television experiments were carried out it was noticed that an equal resolution in the horizontal and vertical directions was obtained with a smaller bandwidth than that given by (5), owing to the discontinuous reproduction in the vertical direction. This was taken into account by including a so called Kell factor in formulae (5). However a rather large spread is present in values reported by various investigators. As already noted, investigations into the relationship between standard and picture quality still continue. Some years ago Jesty made a very interesting attempt to determine picture quality on the basis of the voluntarily preferred viewing distance. The relation between picture quality in the vertical and horizontal directions was determined in this way ^{6,7)}.

*) "Frame" is used, in accordance with American usage, to mean one complete cycle of the scanning process. The usual British nomenclature, in which picture is the equivalent word, is not employed here because this may cause ambiguity due to the second general meaning: the entire visible picture seen by the eye.

REFERENCES

- 1) G. Sziklai, Trans. Inst. Radio Engrs I.T. **2**, 125-128, 1956.
- 2) J. Licklider, Technical Report Acoust. Lab. M.I.T., 1954.
- 3) C. Shannon, Bell Syst. tech. J. **28**, 646, 1948.
- 4) W. Kroebele a.o., Z. angew. Phys. **10**, 320-327, 1958.
- 5) F. Below, Rundfunktechn. Mittn **2**, 184-186, 1958.
- 6) L. Jesty, Wireless World **63**, 304-306, 1957.
- 7) L. Jesty, Proc. Inst. elect. Engrs **105B**, 425-439, 1958.

II. POSSIBILITIES FOR BANDWIDTH COMPRESSION

II.1. Introduction

The search for means of reducing the bandwidth required for the transmission channel is a search for "redundancy" in the conventional signal. Only if a certain amount of redundancy is present in the data given by the signal will it be possible to decrease bandwidth by eliminating these redundant data. The analysis of a conventional television signal reveals three kinds of redundancy. In the first place there is a redundancy based on the television signal statistics. It has often been noticed that a considerable correlation will exist between the signal content of adjacent picture elements and also between the signal content for successive scans of the same picture element. In principle it is possible to make use of this correlation for bandwidth compression. A system based on this principle will reproduce an image which is physically identical to the image reproduced in normal television, but the redundant data are omitted in transmission.

A second kind of redundancy exists which is related to the physiological limitations of visual perception: for instance those on acuity and the speed of response of the eye. Besides the physical boundaries of the picture, these limitations are being exploited in normal television to make the transmission possible. However it is not inconceivable that further use might be made of these physiological thresholds. When this form of redundancy is employed, physically there will be a difference between the image reproduced under this system and a normal television picture. However this will not be perceived by the observer, and so the perceived pictures will be identical. It will be clear that a large subjective element is involved in the judgment of perceptibility. In individuals, moreover, the transition from perceptible to not perceptible is not a sharp one. In consequence the search for a system with reduced bandwidth in which picture quality is maintained also leads to considerations in which a slight loss of quality is set off by a certain saving in bandwidth *).

In the third place a psychological redundancy exists which is related to levels of consciousness. There are special limitations on the apprehension. The amount of information which can be apprehended is much less than can be perceived by the eye ²⁾. For instance the observer's eye is able to perceive ten totally

*) In addition to this it can be noted, that the picture-quality bandwidth-compression problem may also be studied in the opposite way, namely by determining for a certain given narrow channel bandwidth the transmission with optimum picture quality. This will not be considered here. The question has been studied by Deutsch ¹⁾.

different pictures per second but it is impossible to take them in during that brief time. On the other hand there is in the viewer an enormous store of knowledge about objects, situations and changes in situation. When these objects or situations are present in the picture content there is in principle a large degree of redundancy in the data transmitted for reproduction at the receiver. If this form of redundancy were eliminated the mechanism of transmission would have to be modified either in such a way that the physical picture as well as the picture perceived would be different but the ultimate impression would be identical, or in such a way that a lot of knowledge about geometry of objects is put in the receiver so that the data in question can be omitted in the transmission. It will be clear that this subject is very remote from practical engineering.

In the following sections the various forms of redundancy here distinguished will be dealt with in detail.

II.2. Statistical redundancy

In the first place we shall consider statistical redundancy, i.e. that kind of redundancy whose presence does not depend on the properties of visual perception. Of course this definition is not quite accurate. One may assume a certain correlation between picture content in nature and the characteristics of perception. For that reason it is impossible to distinguish sharply between the various forms of redundancy; it is however possible to consider statistical redundancy without analysing the properties of the eye.

Statistical redundancy can be said to be present where a certain correlation occurs between signal values which in principle can be transmitted independently. For instance, the brightness of all picture elements can be given independently but the presence of a certain correlation is admissible. As stated above, this will hold especially for adjacent picture elements and for successive scans of one particular picture element.

Picture elements which are adjacent in the horizontal direction also will be adjacent as video-signal values. It will therefore be easiest to determine and to make use of correlation in this direction. For the vertical direction — in which we may expect correlation similar to that in the horizontal direction — a comparison has to be made between signal values which are separated by one or more line intervals. The correlation between successive scans of the same region extends over time intervals which are still more longer.

The correlation between adjacent picture elements in the horizontal direction will be treated first.

II.2.1. Correlation between adjacent picture elements

The occurrence in most images of large areas with no or only slight variations in brightness has been analysed by several investigators, usually with an eye to bandwidth compression of the video signal. Harrison studied the predictability

of picture-element values¹⁾. He did not refer directly to reduction of bandwidth, but measured the energy reduction in the transmission signal when, instead of the normal signal only the deviation from a predicted signal value (according to some prediction principle) is transmitted. An experimental analysis of the redundancy has been carried out by Kretzmer and Schreiber^{4,5,6)}. Deriugin and Galitskaya have also investigated the correlation of neighbouring picture elements, however, without considering the relation with redundancy and bandwidth compression^{7,8)}. As nothing has been mentioned about the kind of pictures analysed, there figures cannot be used here in determining possible bandwidth compression. The theoretical aspects of the redundancy have been considered by Powers and Staras⁹⁾. Cherry and Gouriet and many others have proposed systems with reduced bandwidth based on the present principle^{10,11,12,13)}. For clearer understanding of the situation, it will be useful to start by studying the implications, in information theory, of correlation between successive picture elements.

According to Shannon the information conveyed by a message which may have k possible contents can be given as

$$-\sum_{j=1}^k B(j) \log B(j) \quad (1)$$

where the probability of the occurrence of the message no. j is given by $B(j)$ ¹⁴⁾. The logarithm is to base 2, as also are the logarithms occurring in the expressions which follow, unless otherwise stated.

If the message can be split up in N parts each having n possible values such that

$$k = n^N$$

formula (1) can also be written as

$$-\sum_{i_1=1}^n \dots \sum_{i_N=1}^n p(i_1 \dots i_N) \log p(i_1 \dots i_N) \quad (2)$$

in which $p(i_1 \dots i_N)$ has been substituted for $B(j)$.

Now, the parts of the message can be identified with the signal values of various picture elements. Hence the information content of one picture element x having n possible values is

$$H(x) = -\sum_{i=1}^n p(i) \log p(i) \quad (3)$$

where $p(i)$ is the probability that the brightness, or signal level, no. i will occur.

For two picture elements the information is

$$H(x,y) = -\sum_{i=1}^n \sum_{j=1}^n p(i,j) \log p(i,j) \quad (4)$$

where $p(i,j)$ is the probability that the first element, x , has a value no. i and the second one, y , a value no. j . When $p(i,j)$ is written

$$p(i,j) = p(i) p_i(j) \quad (5)$$

and the so called "conditional information"

$$H_x(y) = - \sum_{i=1}^n \sum_{j=1}^n p(i,j) \log p_i(j) \quad (6)$$

is introduced, we get

$$H(x,y) = H(x) + H_x(y). \quad (7)$$

Of course it is possible that $p_i(j)$ does not depend on i , that is to say that there is no correlation between the two elements. In that case is $p_i(j) = p(j)$, $H_x(y) = H(x)$ and $H(x,y) = 2H(x)$.

For three elements x , y , and z the information is

$$H(x,y,z) = - \sum_i \sum_j \sum_l p(i,j,l) \log p(i,j,l). \quad (8)$$

Again $p(i,j,l)$ can be written as

$$p(i,j,l) = p(i,j) p_{i,j}(l). \quad (9)$$

If $p_{i,j}(l)$ only depends on j , correlation exists between the elements y and z but there is no correlation between the elements x and z . (Markoff process.) In that case $p_{i,j}(l) = p_j(l)$ and

$$H(x,y,z) = H(x) + 2H_x(y).$$

Similarly, the information for N elements is then

$$H(x) + (N-1)H_x(y). \quad (10)$$

It follows from this that for a sufficiently large number of picture elements the information per element is $H_x(y)$, provided correlation exists only between adjacent elements.

Schreiber has investigated the extent to which the third order correlation is important in comparison with the second order correlation⁵⁾. From these measurements the conclusion may be drawn that the correlations of the third and higher orders are of minor importance. In what follows, only the second order correlation is taken into account and consequently the transmission of picture elements is considered to be a Markoff process in the original sense.

The information per picture element is

$$H_x(y) = - \sum_{i=1}^n \sum_{j=1}^n p(i,j) \log p_i(j). \quad (11)$$

The redundancy per element is

$$H_M - H_x(y) \text{ bits,} \quad (12a)$$

where H_M is the maximum value of $H(x)$, which occurs if $p(i)$ is independent of i and equals $\log n$. This redundancy can be broken into two parts:

$$\{H_M - H(x)\} + \{H(x) - H_x(y)\}. \quad (12b)$$

The first part relates to the statistical distribution of the signal values themselves. The second part relates to the statistical distribution of signal value differences for adjacent picture elements. Only the latter kind of redundancy is considered below. It is true that in principle the redundancy involved in the statistical distribution of brightness values can also be exploited for reducing bandwidth but the difficulties of realization are much greater still, and therefore this possibility is ignored.

We shall use R , the symbol for redundancy, in the restricted sense of

$$R = H(x) - H_x(y) \text{ bits per element.} \quad (12c)$$

As a percentage of the maximum amount of information this is

$$\frac{(H(x) - H_x(y)) \text{ } 100\%}{H_M} \quad (12d)$$

In theory the bandwidth can be reduced to

$$\frac{H_M - \{H(x) - H_x(y)\}}{H_M} 100\% \quad (13)$$

of the normal value. However this cannot be regarded as a value obtainable in practice. A very complicated system and very complex equipment is needed even for far from complete advantage to be taken of the statistical properties of adjacent picture elements. Nevertheless, it is useful to know what is possible in theory. Measurements with a view to evaluating the theoretical maxima have been carried out by Kretzmer and Schreiber. The measurements of Gourié were matched to a particular method of bandwidth compression, his aim being to examine the merits of this special system.

We shall now deal with these measurements in greater detail, describe some measurements of our own and compare the results.

II.2.1.1. *Measurements by Schreiber and Kretzmer*

The investigations carried out by Schreiber are very closely related to the theoretical expression (2) for the information conveyed. Values of $p(i,j)$ and $p(i,j,l)$ are measured in two ways^{5,6}. The first method was based on the measurement of the brightness of an oscillograph screen pattern obtained by deflecting the beam in horizontal direction by the video signal to be measured and in vertical direction by that video signal delayed by a time interval corresponding to the distance between picture elements. A grid of 32×32 squares was placed over the screen pattern, and the light output from each square was

measured. It will be clear that the light output for the square with the coordinates i and j corresponds with $p(i,j)$ for a quantization of the signal range in 32 levels. In the second method two or three video signals only differing in time delay were compared by means of a coincidence circuit embodying a counter. In this way the statistical distribution was determined of the number of instants at which value no. i in the first signal coincided with values no. j and l in the second and third signals respectively. It was possible thus to determine $p(i,j,l)$ and $p(i,j)$. As already mentioned, the third order correlation was found to be relatively small. In this experiment a quantization in 64 levels was applied. When $p(i,j)$ is known $p(i)$ can be determined and also $p_i(j) = p(i,j)/p(i)$. Sufficient data are available then for redundancy to be calculated from (12c). From the results of Schreiber it follows that in normal television pictures R is at least $2^{1/2}$ bits if a quantization in 32 levels is applied. In theory, therefore, 50% of the normal bandwidth would be sufficient.

It is somewhat more difficult to work out the possible bandwidth reduction from the measurements of Kretzmer, because he measured, not statistical distributions, but the autocorrelation present in the signal ⁴). The autocorrelation $n(\tau)$ of the television signal $u(t)$ is expressed as

$$n(\tau) = \frac{1}{T\sigma^2} \int_0^T u(t) u(t-\tau) dt \quad (14)$$

where σ is the r.m.s. value of $u(t)$ and τ is a certain time delay. Since it is a matter of investigating correlation between adjacent picture elements, only one television frame is examined, and hence T is here a frame period. In order to avoid that for large values of τ , in which case autocorrelation can be assumed to be zero, $n(\tau)$ does not vanish due to d.c. component in $u(t)$, this d.c. component is assumed to be eliminated.

From the theory of the Fourier transform it follows that

$$n(\tau) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau} d\omega \quad (15)$$

where $S(\omega)$ is the energy per second and per cycle from the frequency spectrum of $u(t)$.

Under special conditions it is possible to work out a direct relationship between autocorrelation and the parameters in the statistical distributions of the signal. These distributions are the continuous probability distributions $p(u)$ and $p(u_1, u_2)$ where u denotes the signal value in general and u_1 and u_2 represent the signal values of two adjacent picture elements.

When $p(u)$ and $p(u_1, u_2)$ represent Gaussian distributions they can be written

$$p(u) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{u^2}{2\sigma^2}\right) \quad (16)$$

$$p(u_1, u_2) = \frac{\sqrt{\frac{1}{\sigma^4} - \frac{1}{a^4}}}{2\pi} \exp\left(-\frac{u_1^2}{2\sigma^2} - \frac{u_2^2}{2\sigma^2} - \frac{u_1 u_2}{a^2}\right) \quad (17)$$

where

$$\sigma^2 = \int_{-\infty}^{\infty} u^2 p(u) du \quad (18)$$

and

$$a^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_1 u_2 p(u_1, u_2) du_1 du_2. \quad (19)$$

Now, $p(u_1, u_2) du_1 du_2$ denotes how often a value u_1 is followed by a value u_2 . So $\int p(u_1, u_2) du_1 du_2$ is the total integrated time during which a combination of u_1 and u_2 occurs in the signal. Hence

$$Ta^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} u_1 u_2 T p(u_1, u_2) du_1 du_2 = \int_0^T u(t) u(t - \tau_0) d\tau = T \sigma^2 n(\tau_0)$$

or

$$a^2 = \sigma^2 n(\tau_0) \quad (20)$$

where τ_0 is the distance in time between adjacent picture elements.

A connection has now been established between $n(\tau_0)$ and $p(u_1, u_2)$. However we want to express the discrete distributions $p(i)$ and $p(i, j)$, and hence $H(x) - H_x(y)$ and R , in $n(\tau)$. The distributions $p(i)$ and $p(i, j)$ hold for a quantization of the signal. When the intervals of quantization are Δu , $p(i)$ and $p(i, j)$ can be expressed in $p(u)$ and $p(u_1, u_2)$ as

$$p(i) = \int_{\Delta u_i} p(u) du \quad (21)$$

and

$$p(i, j) = \int_{\Delta u_i} \int_{\Delta u_j} p(u_1, u_2) du_1 du_2. \quad (22)$$

The integration is extended over the interval no. i , Δu_i , for the first picture element and over the interval no. j , Δu_j , for the second picture element.

It is not so easy to calculate $p(i)$ and $p(i,j)$, and hence R , for arbitrary values of the intervals Δu_i . Only in the limit $\Delta u \rightarrow 0$ it is rather simple. In that case

$$H(x) = - \int_{-\infty}^{\infty} p(u) \log p(u) du - \lim_{\Delta u \rightarrow 0} \log \Delta u$$

$$H_x(y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(u_1, u_2) \log \frac{p(u_1, u_2)}{p(u_1)} du_1 du_2 - \lim_{\Delta u \rightarrow 0} \log \Delta u$$

and therefore

$$H(x) - H_x(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(u_1, u_2) \log \frac{p(u_1, u_2)}{p(u_1)^2} du_1 du_2. \quad (23)$$

It can be easily shown that by inserting of (16) and (17) into (23), we obtain

$$R = H(x) - H_x(y) = -\frac{1}{2} \log \{1 - n(\tau_0)^2\}. \quad (24)$$

Hence certain assumptions underlay the calculation of redundancy from Kretzmer's data, namely that the distribution is Gaussian and that signal values are quantized in infinitely small steps. It is rather difficult to evaluate the numerical influence of the finite quantization interval on $H(x) - H_x(y)$. However there is close correspondence between the figures found by Kretzmer and those of Schreiber. Ketzmer gives a redundancy figure of at least 3 bits per picture element.

It has to be borne in mind that, from the point of view of bandwidth compression, the figures measured and calculated by Kretzmer and Schreiber represent what is theoretically possible. It will be useful to consider, on the other hand, measurements which are closely related to proposals for practical systems. We carried out such measurements. They are described in the next sub-section.

II.2.1.2. *Measurement of the number of "essential picture elements"*

Two principles have been proposed for relatively simple systems. In the first place it has been proposed to transmit only those picture elements which differ significantly from the previously transmitted element¹²⁾. In the second place variation of scanning speed has been proposed, the speed to depend on the presence of "picture detail". The second principle will be discussed later on. To start with, the first principle will be described.

This principle achieves a bandwidth reduction that only depends on a single aspect of correlation, viz. on how often a difference occurs between adjacent picture elements. Of course the difference has to be quantitatively defined as one exceeding a certain laid down value. To judge the merits of the system, it will suffice to measure the probability of this threshold being exceeded. We have measured this probability figure on the basis of the following reasoning.

A picture element is transmitted only when its signal value differs by more than a fixed amount ΔV with respect to the signal value of the previously transmitted picture element. So in order to determine if a given picture element is redundant or not, it is compared with the last one transmitted, not with the immediately adjacent element. This is necessary because comparison with the adjacent element would not take account of slight changes from element to element extending over a large region, and ultimately giving rise to large picture element differences. When for instance a signal slope occurs corresponding to a difference of $0.2\Delta V$ between adjacent picture elements, this would fall below the threshold value if adjacent picture elements are compared. But over each five picture elements the threshold is exceeded and when the slope continues over the whole picture width the picture brightness varies from full white to black. Of course this cannot be ignored in transmission. Comparison with the previously transmitted picture element will result in the transmission of each fifth element.

For the latter comparison the value of the last transmitted element has to be stored, for a reference, over a period which is a priori unknown. This process is rather complicated. In order to avoid this difficulty yet the difference between adjacent picture elements was measured, but in such a way that a reasonable correspondence could be expected with the result obtained when reference is made to the previously transmitted picture element. This we were able to do taking not merely into account those picture elements for which the difference exceeds ΔV , but also the picture elements for which the difference is below ΔV ; the latter, however, partially, viz. proportional to the difference value. Those picture elements for which the difference exceeds ΔV are taken into account as one picture element. The integrated differences provide a value which, related to ΔV , is a measure of the number of picture elements which have to be transmitted. We shall refer to these picture elements as "essential picture elements".

The method described is quite correct if the slope of signal waveform between two essential elements does not change its sign, as will be clear from fig. 1. In the interval $A - B$ the signal slope is always negative. The integrated difference does in this case equal ΔV , and one picture element is quite rightly taken into account for this interval. However in the interval $B - C$ the integration of the absolute difference value gives a value larger than ΔV and hence more than one picture element is taken into account, whereas according to the definition only one essential picture element is present in the interval. So a certain error will be present in the measurement related to signal variations with an amplitude below ΔV . Of course, the error increases with the frequency of these variations. In order to avoid this error far more complicated circuitry should have to be applied. This has not been done because the measurement as described gives a definite upper limit and, as will be discussed later on, a lower

limit can also be derived with this measurement, so that sufficient data become available after all.

In practice two video signals with a time delay of $0.1 \mu\text{sec}$ were compared by means of the circuit shown in fig. 2. Here De represents a delay line with a delay of $0.1 \mu\text{sec}$, the distance at which picture elements are separated in the 625-lines television system. The delayed signal and the undelayed signal are subtracted one from the other.

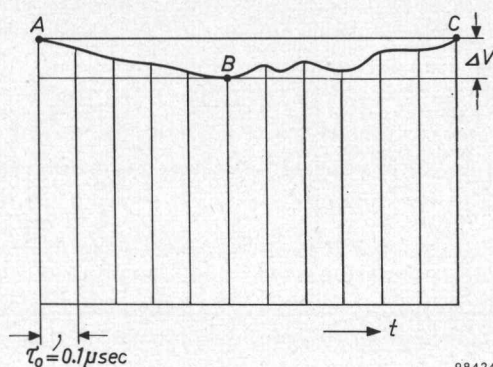


Fig. 1. Video-signal waveform with three "essential picture elements" at A, B and C.

The difference signal is amplified and an amplitude selection is made for levels between zero and ΔV . As the signal level corresponding to zero difference may be affected by tube settings and non-linearities, this level has to be identified by some special arrangement at the point where amplitude selection is carried out. In the blanking intervals the difference signal is certainly zero, and therefore a reference was made to these intervals in order to determine the zero difference level. The voltage level V_0 during blanking periods, was

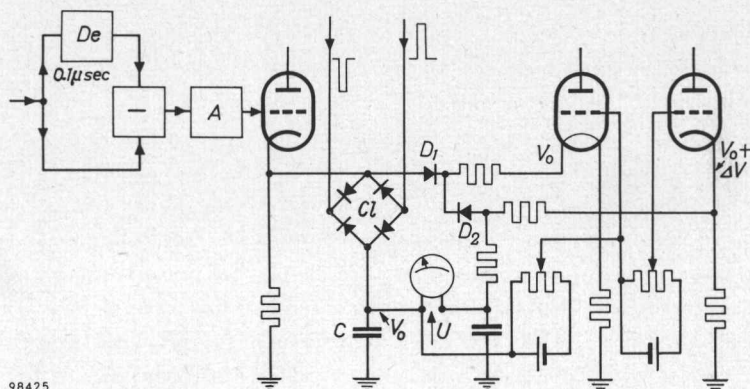


Fig. 2. Measurement of the percentage of "essential picture elements". De = delay of $0.1 \mu\text{sec}$, A = amplifier, Cl = Clamping circuit, V_0 = zero level, ΔV = threshold value.

measured by a clamping circuit CI and fed to a capacitor C . This voltage was fed to a cathode follower with a constant battery voltage in series; the latter was introduced to compensate for the potential difference between grid and cathode. With proper adjustment the zero level V_0 will be present at the cathode and will act as a bias for the diode D_1 . The second diode D_2 is biased with a voltage which is equal to $V_0 + \Delta V$.

The amplitude selected difference signal is present at the anode of D_2 . The mean value U is measured by means of an RC coupling. Of course, V_0 has to be subtracted, and therefore the second terminal of the meter is connected to V_0 . It will be clear that only differences of one polarity are taken into account. However, an amplitude selection between 0 and $-\Delta V$ has been omitted because it was assumed that, in general, both selections will show the same amount. This can be easily verified by altering the polarity of the input signal.

The video signals were derived from diapositives by means of a flying-spot scanner¹⁵⁾. A certain inaccuracy in the measurement is caused by the ever present parasitic signal components caused by the noise of the photomultiplier as well as the phosphor structure of the flying-spot tube. This interference contributes to the measured values. Of course it is possible to measure the interference alone, by omitting the diapositive and by adjusting the brightness level to a mean value. However there is some doubt about the way in which the two results have to be combined in order to compensate for the interference effect. This is only possible if the statistical properties of both signals are known.

For Gaussian distributions the problem is fairly simple because in such a case the distribution character is not changed by linear operation or by the addition of an uncorrelated second Gaussian distribution. Hence, when the relation between the value U and the r.m.s. value σ_D of the signal fed to the amplitude selector is known, from U_{S+N} (measured value for picture signal and noise) and U_N (measured value for noise only), the value of U_S (picture signal only) can be found for Gaussian distributions from the formula $\sigma_{D(S+N)}^2 = \sigma_{D(S)}^2 + \sigma_{D(N)}^2$.

The relation between U and the r.m.s. value σ_D of the difference signal can be written as

$$U = \int_0^{\Delta V} \frac{u}{\sigma_D \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_D^2}} du + \int_{\Delta V}^{\infty} \frac{\Delta V}{\sigma_D \sqrt{2\pi}} e^{-\frac{u^2}{2\sigma_D^2}} du. \quad (25)$$

The value of $U/\Delta V$ as a function of $\sigma_D/\Delta V$, as expressed by (25), is plotted in fig. 3. It can be seen from this graph that for small values of $\sigma_D/\Delta V$ an almost linear relationship exists. For values of $\sigma_D/\Delta V$ in this region is

$$U_S^2 = U_{S+N}^2 - U_N^2. \quad (26)$$

The question arises as to what extent the difference signal shows a Gaussian distribution and as to the permissibility of subtracting the squares of the r.m.s. values (the noise component can be supposed to be of a Gaussian character.)

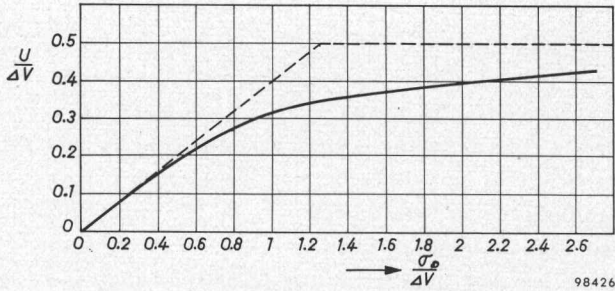


Fig. 3. The relation between the percentage of "essential picture elements", $(2U/\Delta V)$ 100% and the r.m.s. value of the difference signal, σ_D , for a Gaussian distribution.

An answer on the first question can be found by analysing the distribution of the difference signal with noise. This distribution can be determined by varying the threshold ΔV in the amplitude-selection process. If U_{S+N} is known as a function of ΔV the probability distribution of the difference signal can be found by differentiating twice U_{S+N} as a function of ΔV . The curve of $U_{S+N} = f(\Delta V)$ for a given picture appears in fig. 4, the derivatives being represented in fig. 5. It can be seen that the shape is rather different from that of a Gaussian curve. Similar curves were found for other pictures. A fairly good qualitative correspondence exists with the results of Kretzmer and Schreiber, insofar as can be ascertained from the rather spare data given.

Since, as is evident, the statistical distribution in the picture difference signal differs considerably from a Gaussian distribution, the quadratic combination of the values obtained will not be permissible without further data. In order to examine this question the effect of an increase of noise above the normal level was investigated. For several noise levels U_S was calculated as if the dis-

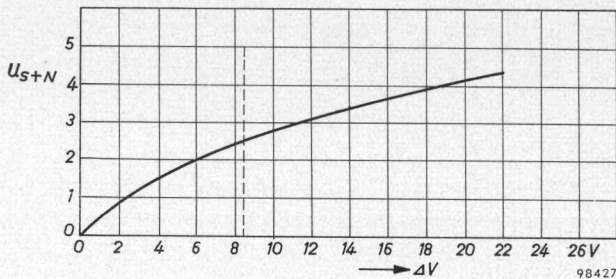


Fig. 4. The measured value U_{S+N} as a function of the threshold ΔV , for the picture of fig. 6. (1/32 of the peak-value of the original video signal corresponds to 8.4 volts.)

tributions were Gaussian in both picture difference signal and noise for several noise levels. As only slight differences were found, we concluded that the above method of allowing for noise gave a fairly good approximation despite the non-Gaussian distribution.

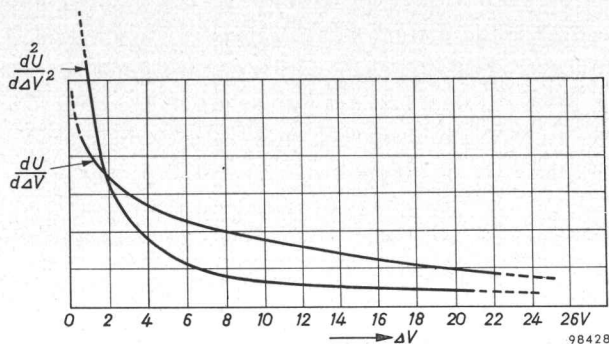


Fig. 5. The derivatives $U'(\Delta V)$ and $U''(\Delta V)$ of fig. 4 as a function of the threshold ΔV . (Relative scale for U' and U'' .)

The value of U was measured for a large number of pictures. The threshold ΔV was fixed at $1/32$ of the peak value of the original video signal. Of course, a large spread was found. Measured values of U_{S+N} varied from $0.12 \Delta V$ to $0.3 \Delta V$ for various pictures. For pictures with little detail, so for low values of U_{S+N} , it is difficult to make a statement about the value of U_S because in such pictures the difference between U_{S+N} and U_N (U_N was about $0.1 \Delta V$) approaches the magnitude of the error in U_N . The latter error is mainly determined by the possibility of getting close correspondence between noise conditions when U_N and when U_{S+N} is measured. Because of this, and because a conclusion about bandwidth compression has to be based on maximum values of U_S , only pictures with a great deal of detail are considered. Such pictures are represented in fig. 6, 7 and 8 (on p. 119 and 120); the data about these pictures are listed below.

	U_{S+N}	U_N	U_S	ΔV	percentage essential elements	
					upper limit	lower limit
fig. 6	2.53	0.82	2.4	8.4	57%	42%
fig. 7	1.9	0.82	1.72	8.4	41%	29%
fig. 8	2.15	0.82	1.9	8.4	45%	30%

The "upper limit" follows from

$$2 \frac{U_S}{\Delta V} 100\%.$$

The "lower limit" calls for further explanation. As already stated, measurement necessarily involves a certain error which is related to small signal variations within the amplitude interval ΔV (fig. 1). The value found is too high, and therefore $2U_S/\Delta V \cdot 100\%$ is denoted as an upper limit. However, a lower limit can also be found, in which these small signal variations are not measured, but in which other picture element differences are ignored, though they ought to be taken into account. Such a value is obtained if only differences exceeding ΔV are considered, so the percentage of time is measured in which the difference signal exceeds ΔV . This can be easily derived from fig. 5 as this value is equal to $U'(\Delta V)$. It is given in the last column, as lower limit. In our opinion there is no point in making further attempts to determine U_S more accurately. The percentages quite clearly show that in certain pictures the number of essential picture elements is too high for a system transmitting only these picture elements to be successful in practice. It appears that this proportion can be as high as 50%; consequently half the normal bandwidth will still be required. Any further analysis is therefore mainly of theoretical interest. An exception can be made for communication systems handling special picture material. For reasons of completeness in the next sub-section an additional discussion is given.

II.2.1.3. Discussion

In the first place the measurement will be compared with the results obtained by other investigators. The lower limit for the percentage of essential picture elements given in II.2.1.2 is directly related to the statistical distribution measured by Schreiber. It is equal to

$$\left\{ 1 - \sum_{i=1}^n p(i) p_t(i) \right\} 100\%. \quad (27)$$

It is much more difficult to write a similar expression for the upper limit U_S . A comparison with the figures of Kretzmer can be made if the picture signal is assumed to have a Gaussian distribution, because in that case U_S can be easily related to the auto-correlation function as will be described below.

The difference signal we used can be described as the operator

$$\{ 1 - \exp(-j\omega\tau_0) \} \quad (28)$$

applied to the frequency band of $u(t)$, where $u(t)$ is the original video signal and τ_0 the picture element interval. If $S(\omega)$ represents the energy spectrum

of $u(t)$, the energy spectrum of the difference signal can be given by

$$|1 - \exp(-j\omega\tau_0)|^2 S(\omega) = \{2 - \exp(j\omega\tau_0) - \exp(-j\omega\tau_0)\} S(\omega).$$

Hence the m.s. value of the difference signal is

$$\sigma_D^2 = \frac{1}{2\pi} \int_{-\infty}^{\infty} (2 - e^{j\omega\tau_0} - e^{-j\omega\tau_0}) S(\omega) d\omega. \quad (29)$$

Now according to (15)

$$n(\tau_0) = \frac{1}{2\pi\sigma^2} \int_{-\infty}^{\infty} S(\omega) e^{j\omega\tau_0} d\omega$$

so

$$\sigma_D^2 = \sigma^2 \{2 - n(\tau_0) - n(-\tau_0)\}. \quad (30)$$

In general the autocorrelation function will be symmetrical with respect to $\tau = 0$. In that case σ_D^2 can be given by the expression

$$\sigma_D^2 = 2\sigma^2 \{1 - n(\tau_0)\}. \quad (31)$$

For small values of $\sigma_D/\Delta V$, U is proportional to σ_D (see fig. 3). In this range

$$U = \sigma_D/\sqrt{2\pi}. \quad (32)$$

The percentage of redundant elements is

$$(1 - 2U/\Delta V) 100\% \quad (33)$$

or according to (31) and (32),

$$\left[1 - \frac{2\sigma}{\Delta V} \sqrt{\frac{1}{\pi} \{1 - n(\tau_0)\}}\right] 100\%. \quad (34)$$

In bits per picture element it is

$$R = \left[1 - \frac{2\sigma}{\Delta V} \sqrt{\frac{1}{\pi} \{1 - n(\tau_0)\}}\right] \log 32 \quad (35)$$

where σ is the r.m.s. value of the original video signal.

The redundancy as measured by Kretzmer was — see formula (24) —

$$R = -\frac{1}{2} \log \{1 - n(\tau_0)^2\}.$$

It has to be borne in mind that in our measurements only a single aspect of correlation has been considered and a lowering of quality is involved because

only 32 signal levels can be transmitted. Moreover, nothing has been done to fix the position of the picture elements that are transmitted. Besides being applied with signal values for the essential picture elements, the receiver has to be informed about their positions.

The most obvious way of conveying luminance and position information about the essential picture elements is to transmit two series of equidistant signal values, one series carrying luminance data, the other series carrying data about position. The bandwidth required for this purpose would be twice as large as follows from the number of picture elements. The transmission of position information then would not be as efficient as it might be, but could only be improved by some complicated encoding process ¹²⁾.

Alternatively, it is conceivable that combined data about position and luminance could be transmitted on the analogy of telephone communication between a large number of subscribers. Theoretically, the telephone system has to provide for a very large number of conversations taking place simultaneously. In practice only a few connections are utilized, because the mean number of telephone calls per unit of time is relatively small. Likewise only a part of the total number of picture elements is essential, hence calls for transmission. Extending the analogy, a television transmission may be imagined consisting of a certain number of narrow-band channels. The essential elements would be distributed over these channels in the sequence of occurrence. Of course, per channel the time interval between picture elements would have to be at least as large as would be appropriate to the channel bandwidth. It will be clear that in this type of transmission system errors might occur if occasionally a large number of essential picture elements arose within a short time. On the other hand it has to be recognized as a good feature of this system that it would not require a special channel for the position signal with the same bandwidth as the brightness channel.

No matter how transmission is achieved, the transmission signal will always carry position data. Consequently any interference in transmission will affect the brightness in the reproduced picture as well as the picture geometry. Picture quality can be expected to be very sensitive for this latter kind of distortion, and the nuisance of "noise in geometry" will be more than that of "noise in brightness" for similar signal-to-noise conditions. So special requirements hold for the transmission, which in fact means another increase of bandwidth. The question will be examined in more detail in sub-section II.2.1.5.

II.2.1.4. *Systems proposed by Cherry and Gouriet*

From the foregoing considerations the conclusion can be drawn that correlation between adjacent picture elements does not provide very promising possibilities for bandwidth compression; this is more or less in accordance with

the experiences of Kretzmer and Schreiber. However, the work of Gouriet and Cherry has not been given attention so far, and deserves further comment in view of the figures of between 2% and 10% that it gives for possible bandwidth compression ^{10,16}).

These figures are related to a measurement of so-called "picture detail". Picture detail has been defined as the mean absolute value of the slope in the video signal. The percentages mentioned were arrived at by comparing the actual picture detail with the maximum possible detail, which occurs when the video signal is a sine wave of maximum amplitude and of a frequency which equals the upper frequency of the television signal bandwidth. The percentage obtained equals the possible bandwidth reduction in a system where the scanning speed varies in inverse proportion to the signal slope ^{10,13}). It is better to say that the time needed for a complete scan is reduced to that percentage. This statement is more exact because the possible reduction in bandwidth does not follow automatically from the reduction in time. This is because scanning with variable speed is in fact a non-linear process, and hence frequencies will be introduced that are beyond the normal frequency range. On the other hand these signal components can be expected to be fairly small.

A more serious objection is that the scanning speed is not matched to the instantaneous "frequency content" but to the instantaneous signal slope. In consequence, though the resulting transmission channel has sufficient bandwidth for extreme swings (from black to white and vice versa), the available bandwidth is inadequate for smaller excursions with the same (original) frequency content. Nor is picture detail any measure of information content as defined in information theory. For special statistical distributions a proportionality exists between picture detail and $H_x(y)$, but the two quantities are not numerically the same ¹⁰).

These objections do not apply to another system proposed by Gouriet and Cherry ¹⁰). Here only two values of scanning speed occur, a high speed if no significant brightness change along the scanning line is present and a low speed if there is a change in brightness. In the latter system, however, the bandwidth reduction is not dependent on the amount of "picture detail"; the results to be expected are the values found in our measurements. Also, in the systems of Gouriet and Cherry certain position data will be present in the transmitted signal, though no separate channel is needed for this information. We will now pay attention to the effect of noise in the transmission of position data.

II.2.1.5. *The effect of noise in the transmission of position data*

It has already been noted that the influence of a certain distortion in the position data upon the ultimate picture quality has an important bearing on the overall usefulness of the system. This holds for all systems based on the

elimination of statistical redundancy. In order to get a better insight in this problem we carried out some experiments. For this purpose a monitor was provided with an extra deflection coil through which "noise currents" were passed. The arrangement is shown in fig. 9.

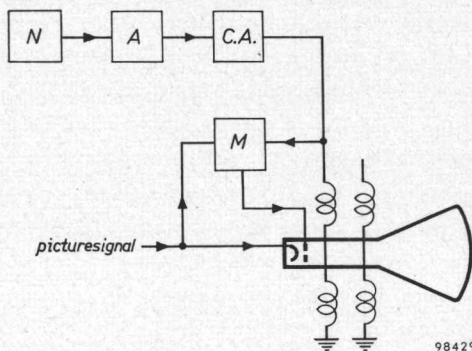


Fig. 9. Arrangement for investigating the influence of noise in the transmission of position data. N = noise source, A = amplifier, $C.A.$ = correction amplifier, M = modulator.

The amplified voltage from a noise source N (noise diode) is fed to a correction amplifier $C.A.$ where output goes to the deflection coil. The effect of the correction is to get for the relation coil current versus input voltage a characteristic which is flat up to about 2 Mc/s. Picture reproduction is now accompanied by a noise effect that dislocates the elements of the picture. The deflection coil was positioned in such a way that the displacement of the picture elements was in the horizontal direction.

Besides the changes in location, another noise effect is present, inherent variations in scanning speed causing fluctuations in brightness. It may be supposed that a practical transmission system will be designed in such a way as to compensate brightness variations due to the process of locating the picture elements. These variations are in inverse proportion to the scanning speed, so to the derivative of the position signal. If compensation takes place at the receiver side by the modulation of luminance signal proportional to the derivative of the position signal, the luminance effect of the "noise in the position signal" will naturally be compensated at the same time. If pre-correction of the luminance signal at the transmitter is carried out, the luminance effect of transmission noise in the location will not be eliminated.

In our experiments we wanted only the effect of position noise to consider. Therefore measures had to be taken to compensate the luminance effect. The correction voltage needed has to correspond with the instantaneous scanning speed, as the reproduced brightness is inversely proportional to this. This vol-

tage is found across the deflection coil if L/R for this coil is sufficiently large. See fig. 9.

With the aid of ten observers the level of perceptibility and the level of annoyance were determined for a viewing distance equal to five times the picture height. The magnitude of position noise was measured by training a microscope on a very narrow white line in the screen image, which was otherwise black. It is rather difficult to determine accurately a r.m.s. value, a mean absolute value or a peak-to-peak value of the deviation in position. The least difficult cause is to determine a peak value. The position noise appeared to be visible for a peak-to-peak value of about one and a quarter the picture element separation as present in a 625-lines standard television picture.

In the annoyance test the picture was compared with another picture in which normal noise was present, the ratio between peak signal voltage and r.m.s. noise voltage being 30 dB. This is the value generally accepted as the threshold of acceptability. The position noise was adjusted by the observer to a point such that both pictures were judged to be of equivalent quality. This was found to be the case for a peak-to-peak value of position noise of about two and a half times the picture element separation.

In a certain sense the appearance of normal brightness noise and position noise are complementary. The brightness noise is most visible in large areas and less in picture detail, whereas position noise does not affect the large area but causes irregularity in the outlines which is very disturbing. From the tolerances measured, it follows that the required signal-to-noise ratio for the position signal is much higher than for the luminance signal if the beginning of the scanning is taken as reference. Of course, a more complicated encoding might be devised, giving better signal-to-noise conditions¹²⁾. However, apart from the obvious disadvantage of a more complicated system, there would be a probability for very large errors in location due to noise. This is inherent to a more complicated encoding.

II.2.2. *Correlation between successive frames*

The correlation between the brightness values of a picture element at the successive instants of scanning of that element is related obviously to the correspondence between the successive "shots" in which the television system conveys information from the scene being broadcast. It hardly need to be stressed that this correspondence is very considerable in many cases. Similarity between successive frames depends on the amount of action in the scene as well as on the handling of the camera. Apart from changes caused by a change of camera or a change of cameralens the differences can be said to be relatively small, because in general both the scene and the position of the camera change very slowly in comparison with the frame rate. However, this does not necessarily imply that the correlation mentioned above will be large. One may think

on a very slow camera movement — for instance when the camera is “panned” around — in a stationary scene. The picture content of two successive frames will be almost identical but the correspondence of picture elements from the scene and elements of the television picture varies, and therefore the correlation between successive luminance values of a certain element of the television image may be small.

The considerable correspondence between successive frames had led also to several proposals for reducing the bandwidth required in a television channel. Among others, Schröter proposed a system in which the scanning speed would be high for picture areas with a stationary content, and slow for picture areas whose content was changing in time ¹⁷⁾. Time is saved by the high scanning speed, so that the more time is available for scanning changing areas; the video frequencies generated are therefore lower, they can be transmitted within a narrower band. Special solutions are also given by Schröter for the problems which would arise in his system in consequence of the variable frame period, the accumulation of interference, the need to synchronize scanning at transmitter and receiver, and the transmission of sudden large changes in picture content. One special device required in this system — which also will be considered in III.2 — is a memory device for storing the television signal so that succeeding frames can be compared.

An investigation into correlation between successive frames has been carried out by Kretzmer ⁴⁾. In his article this is described rather briefly. The experiments are similar to his experiments mentioned in II.2.1.1. Instead of autocorrelation the crosscorrelation of two successive frames is measured. The redundancy is calculated from correlation in a way similar as was described in II.2.1.1. A value of 1 bit per element has been found, this corresponding to 20% redundancy for quantization in 32 steps.

We for our part have not performed any experiments relating to this type of redundancy. However the results given in sub-section II.2.1.1 may be used to give some insight into it. Suppose that a part of the picture content moves and the rest can be considered as a stationary background. Now, if the speed of horizontal movement corresponds to exactly one picture element per frame interval, then for the moving part the difference between successive frames is similar to the difference signal as used in II.2.1.1 for determining the essential picture elements. When a shift over one picture element occurs for the whole picture — as may be caused by the camera moving — the values of II.2.1.1 for the percentage of essential picture elements and for the possible bandwidth reduction, are directly applicable here. This makes it clear that considerable correspondance between successive frames may be attended by a rather small possible bandwidth reduction if the latter is based on fixed picture elements.

The speed of movement just assumed is very low. For a speed higher than

one picture element per frame interval almost all picture elements of the moving part would have to be taken into account, owing to the rapidly decreasing degree of correlation. This simple reasoning brings us to the conclusion that camera handling reduces correlation of the kind considered here to a very low level. The correlation will be less reduced by movement occurring in the scene itself. The image mostly changes only partially.

In general it can be said that time intervals with large and small amounts of correlation will occur alternately. It would seem to be no easy matter to devise a practical coding system which would cope with these heavy fluctuations in the information rate.

II.3. Physiological limitations

Although it is not always possible to distinguish strictly between physiological limitations and psychological limitations in visual perception, this classification can be applied to most phenomena. Roughly it can be said that physiological limitations refer to the eye and the transmission system coupled to it, and psychological limitations refer to the processes of consciousness taking place in the brain. In this section we shall consider the physiological limitations; an analysis will be given of resolving power, persistence of vision, ability to distinguish differences in brightness and the ability to distinguish colour differences.

II.3.1. The resolving power of the eye

It is well known that the nominal resolving power of the eye only applies to a very small region in the centre of the fovea. Towards the edges of the retina the resolving power decreases rapidly. Figure 10 shows this decrease as measured by Wertheim¹⁸). (For horizontal displacement, similar curves hold for other directions.) So when the television picture definition corresponds to the maximum resolving power in the whole picture area the information available is considerably more than can be perceived by the eye. For our present purpose

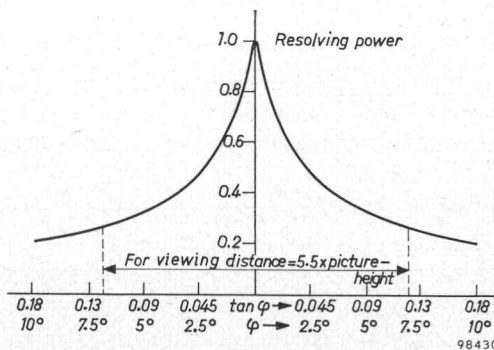


Fig. 10. The place-dependent resolving power of the retina.

it is of interest to examine the situation quantitatively. We will therefore consider a television picture with a place-dependent sharpness which is matched to the local resolving power of the retina. First, of course, the position of the observer has to be chosen. We assumed the observer to be at a distance of $5^{1/2}$ times the picture height, staring at the picture centre. Accordingly the picture is seen within an angle of about 15° . This distance has been chosen because in this position the relative picture element size of a standard 625-lines television picture corresponds to the maximum resolving power of the eye. (Bandwidth is 5 Mc/s, the resolving power is $1.5'$.) For the sake of simplicity the characteristics of the eye are assumed to be the same at equal distances from its geometrical centre, the resolving power falling off radially as in fig. 10.

In principle the number of picture elements required per cm^2 at a certain place in the television picture is proportional to the square of the corresponding resolving power of the eye S . However, realization would imply a varying number of lines per cm which seems out of the question. Therefore we will compute the number of picture elements of a television picture where the number of lines remains constant but only the horizontal dimension of picture elements changes, in accordance with resolution. Using fig. 10, we find a figure equal to a quarter of the normal number. The bandwidth required is thus a quarter of the normal bandwidth. This bandwidth compression is realized if the scanning speed varies in inverse proportion to S and the normal frame period is maintained.

We are not of course asserting that it is possible to reduce bandwidth in practice. Obviously, the eye movements cause insurmountable difficulties.

In connection with our subject the following data about eye movements are of interest¹⁹. When a stationary part of the picture is observed the eye is fixed during a certain time interval on a certain point in the picture and then jumps rapidly to another place for another pause. The fixation intervals vary from one tenth of a second to several seconds. The speed in the transition movement varies between 100° and 200° per second for small distances and between 200° and 500° for larger distances. Slow movements of the eye during which information is absorbed occur if a series of stationary objects are looked at successively. Movement is not continuous but has a sawtooth form; here too, therefore, it consists of a series of fixation intervals interrupted by rapid transitions. This kind of eye movement is the usual one in reading.

When a moving object in the picture is observed, two kinds of slow eye movements may occur. Firstly, there may be smooth movement coupled to the moving object. This can be said to be a fixation pause with respect to the moving part in the picture. In the second place a sawtooth-like motion may take place; this occurs when a series of moving objects comes along and a short period of watching is interrupted by a jump back to the next object (nystagmus). What is said here, of course, only concerns the main features of voluntary eye movement. If we were treating the subject in more detail we should have to consider (1) the converging approach to a fixation point in jumping from one point to the other (a certain "overshoot" is always present) and (2) the very small eye movements still present during fixation (periodic excursions of $4'$ at intervals of 1 to 2.5 seconds).

The pattern of the eye movements naturally depends on picture content and is very individual. In order to test the situation practically, an investigation was made about two aspects. In the first place the impression obtained from

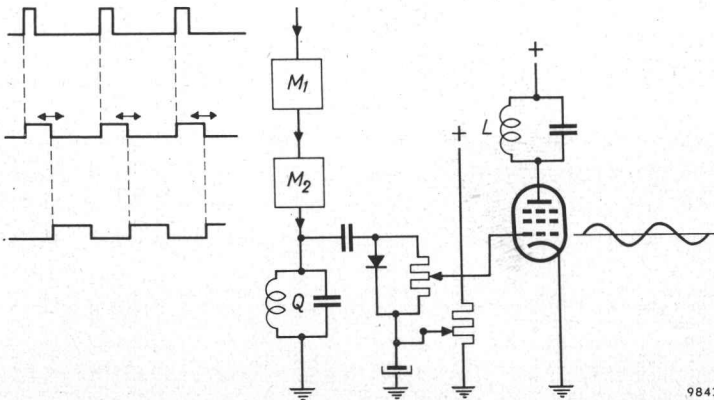
a picture with place-dependent sharpness was examined. In the second place we studied the reaction of an observer to a picture with a sharpness centre shifting in accordance with a pattern which could be assumed to match the events in the picture in a more or less logical way. In these experiments, place-dependent sharpness was introduced into the television picture not by varying the scanning speed, but with the aid of dynamic synchronized defocussing.

Defocussing can be effected quite simply by adding components of line and field frequency to the d.c. current through the focussing coil. As the high inductance of this coil make it difficult to get a component of line frequency with a reasonably high amplitude, in practice an extra focussing coil (or rather a defocussing coil) is needed. The component of field frequency can be fed through the normal coil without any difficulty. In fact the defocussing is required to take place in exact conformity with the curve in fig. 10. Of course, for very close correspondence, rather complicated waveforms have to be employed. In our experiments, however, only sine waves were used. As in a first approximation sharpness can be assumed to be inversely proportional to spot diameter, the sharpness varies according to

$$\frac{1}{1 + \alpha(1 - \cos x)} \quad (36)$$

with maximum definition for $x = 0$. This gives a reasonable analogy with the curve of figure 10. Figure 11 shows the arrangement employed.

A monostable multivibrator M_1 is synchronized by field synchronizing pulses. The leading edges of the multivibrator waveform coincide with the field synchronizing pulses, whereas the position of the trailing edges is determined by the RC product of the multivibrator. This RC value is variable. The trailing



98431

Fig. 11. Arrangement for obtaining a television picture in which the definition is matched to the place-dependent resolving power of the retina. M_1 and M_2 are monostable multivibrators, Q = high- Q circuit, L = focus coil.

edges trigger a second monostable multivibrator M_2 which delivers a square wave. By means of a high- Q circuit the square wave is transformed into a sine wave. The frequency is of course 50 c/s and the phase can be controlled by the setting of M_1 . The sine wave is clamped to a certain potential by a diode and can be varied in amplitude by a potentiometer. It is fed to a pentode having the focus coil L in its anode circuit. By varying the clamping level the equipment can be adjusted in such a way that maximum sharpness occurs at the peaks of the sine wave. Defocussing in line frequency was obtained in a similar way. However, here a separate circuit corresponding to Q in fig. 11 was not needed, because the added focus coil itself could be used to produce a sine wave. In order to indicate the location of the sharpness centre in the picture, a marking pulse was added to the video signal.

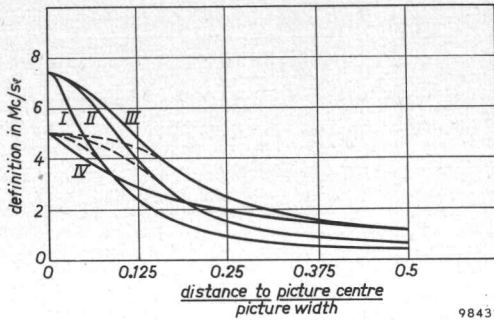
With the aid of this equipment the following experiment was carried out. An observer at a distance equal to $5\frac{1}{2}$ times the picture height was asked to stare at the point in the picture centre with maximum definition, and to say when any unsharpness was perceived. Starting with normal definition the unsharpness towards the edges was increased very slowly. The pictures were derived by means of a flying-spot scanner from diapositives like figs. 6, 7, 8, on p. 119 and 120. Of course there is a certain spread in the results, obtained from different observers as well as those in respect of different pictures. However this spread was relatively small, as can be seen from the data given below.

As regards the results obtained in different pictures, the obvious effect occurred that for much detail, especially with a regular pattern, the unsharpness was noticed more promptly. This is illustrated by fig. 12 and fig. 13 (see p. 120 and 121), being photographs of screen images into which equal amounts of unsharpness have been introduced. They should be compared, the observer fixing his eyes on the centre of the picture and maintaining a suitable viewing distance. ($5\frac{1}{2}$ times the picture height.)

From the results obtained four types of adjustment were chosen for which the corresponding decrease in definition over the picture area was determined quantitatively. This was carried out by measuring the width of the television lines. The result is given in fig. 14 (for the horizontal direction, in the vertical direction the curve was almost identical; in other directions some deviation occurred which could be tolerated). Curve I represents the lower limit of perceptibility. Such a degree of defocussing was detected on 90% of the occasions. Curve II represents the upper limit. This degree of unsharpness is only noticed in very special test patterns. For the latter kind of patterns, defocussing had to be in accordance with curve III if it was not to be detected.

In fig. 14 the definition of the screen image is expressed in Mc/s. It will be noticed that the maximum definition considerably exceeds 5 Mc/s. This is due to the fact that the measurements were carried out by measuring the width of the television lines. Of course this line width is not influenced by frequency

limitation of the video signal. For measurements of vertical lines and a bandwidth of 5 Mc/s the definition curve will have a form like that of the dotted line in fig. 14. Figure 14 also shows the resolution curve of fig. 10 matched to the circumstances of the experiment. (Curve IV.) That is to say, the optimum resolving power has been assumed to correspond with 5 Mc/s.



98432

Fig. 14. Picture definition as a function of distance to the picture centre for three types of defocussing adjustment (I, II, III). The curve of fig. 10 is given as a reference (IV).

From fig. 14 it may be concluded that a satisfactory degree of correspondence exists between tolerable unsharpness in the test pattern and the resolving power of the retina as referred to in the literature. Slightly more unsharpness is tolerable in normal screen images. As the fixation intervals exceed the frame interval considerably, about 75% of the information offered by the television picture is in fact redundant, owing to the limited resolving power of the eye.

Though it is difficult to imagine any practical application of this knowledge, a further experiment was carried out with a moving sharpness centre in order to investigate the extent to which its movements could be linked to the eye movements of the observer. For this purpose a video signal was derived from a film and the position of the sharpness centre was varied in accordance with the viewing pattern of the man handling the equipment. As before the observer was located at a viewing distance of $5\frac{1}{2}$ times the picture height and kept his eye on an electronic marker which shifted about within the picture area. As the manner of controlling the sharpness centre was rather primitive, (two knobs being manipulated, one for horizontal and the other for vertical displacement) it was impossible to establish a very close correspondence between the observation of the operator and the location of the sharpness centre. Therefore not very much can be said about what might ultimately be possible in this direction. On the other hand however the results were of so little interest that no further elaboration of the experiment was made.

During observation of the television picture, just described, large parts of what is shown can be accepted but now and then annoying effects occur, especially when two or more objects of interest are present at the same time.

Normally rapid eye movements of the observer from one object to the other occur in that case, but a particular pattern cannot be predicted.

Finally, we should like to note the following conclusion. In principle the resolution power of the eye allows a reduction in bandwidth of about 75%. Any reduction in bandwidth, however, would mean imposing a "line of sight" on the viewer, and the greater the reduction the more directional viewing would have to be. Restricted application of this principle by halving the bandwidth, for example, would not appear to yield any great advantage to compensate for the rather complicated technique which would be required. (variation of scanning speed and control of the sharpness centre position by the transmitted signal.)

II.3.2. *Persistence of vision*

In this sub-section we shall consider a second characteristic of visual perception, namely persistence of vision. In this two aspects can be distinguished: on the one side the ability of the eye to perceive rapid variations in brightness and on the other hand its ability to distinguish a rapid sequence of stationary pictures from a real natural continuous motion. As is well known, in both respects the eye has its limitations.

Investigations into the perception of rapid variations in brightness, i.e. the flicker effect, have been carried out by De Lange and by Haantjes and De Vrijer^{20,21}). The first-named investigator studied the phenomena in a general way and made also an attempt to derive from his measurements a model for the process by which visual impressions are transmitted to the brain. Haantjes and De Vrijer studied the flicker effect in relation to television in particular.

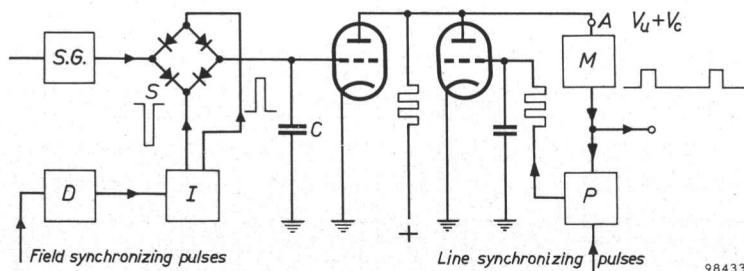
Fewer data are available about awareness of discontinuity in motion. Though from television and film practice it is known that 25 situations per second give a satisfactory result and that 16 situations per second, as in 8 mm film, may also be adequate, the limit as dependent on the various parameters is not precisely known. In television, however, it is important to know how a television picture is appreciated when the number of frames is intentionally dropped below the threshold for continuity. This is of interest when it is desired to transmit over large distances and to cut down the band occupied by the signal by simply decreasing the number of frames per second. For these reasons some experiments were carried out on the perception of discontinuity; they are described below.

II.3.2.1. *Discontinuity in motion*

In our experiments on the perceptibility of discontinuity in motion the number of situations reproduced per second could be varied, the normal field rate necessary to avoid flicker being maintained. In a first experiment an "elementary" picture content was used, namely a vertical bar moving in horizontal direction. This picture signal can be generated and varied in a simple

way, but only data about the extreme phenomena are obtained. Another experiment was conducted with a complex picture content. Such an experiment needs elaborate technical equipment but gives more data about overall picture quality.

Figure 15 shows the circuit for the first experiment. The picture signal has a rectangular waveform of line frequency and is supplied by the multivibrator M . The frequency of M depends on the voltage at A because the discharge of one of the capacitors is controlled by this voltage. A frequency equal to the line frequency is obtained for a special value of the voltage at A . This voltage consists of two components, voltage V_u and the control voltage V_c . Voltage V_c is derived from a phase discriminator P in which the multivibrator frequency is compared with line-synchronization pulses. The phase discriminator is a bistable multivibrator which can be triggered into one state by pulses derived from M and into the other by line-synchronization pulses. The output voltage has a rectangular waveform in which pulse width depends on the phase difference between the line-sync pulses and the picture signal as supplied by the multivibrator. The mean output voltage of the discriminator, which is likewise dependent on this phase difference, is applied to the grid of a triode; V_c is a component of a voltage arising in the anode circuit of the triode. The result is that the multivibrator M is synchronized to line frequency in a particular phase. The component V_u is supplied by a second triode connected in parallel with the first mentioned one.



98433

Fig. 15. Arrangement for obtaining a test signal corresponding to a vertical bar moving horizontally with a prescribed number of situations per sec. M = multivibrator, P = phase discriminator, D = frequency divider, $S.G.$ = saw-tooth generator, S = sampling circuit, I = pulse shaper.

Now when V_u is varied a new value of V_c will arise such that $V_c + V_u$ is kept constant. Consequently, the picture signal will undergo a further phase shift with respect to the line pulses. In this way the phase of the picture signal can be varied with respect to line synchronization; in other words, the position of the vertical bar in the picture can be changed. So when n situations of the bar are wanted per second, the voltage V_u has to be varied by n steps per second.

In order to obtain such a step shape for V_u a sawtooth voltage derived from the sawtooth generator $S.G.$ was used combined with a sampling circuit S . Sampling takes place n times per second, and by this means the sawtooth voltage is applied to capacitor C for very brief periods. The RC values are chosen in such a way that the capacitor C is charged very rapidly up to the sawtooth voltage and that this voltage is maintained between successive samples. In this way the desired sawtooth step waveform becomes available from C . The sawtooth is of constant amplitude which value corresponds to a shift of the moving bar over the whole picture width and of variable frequency in order to obtain various speeds for the moving bar. Sampling pulses are derived from field synchronizing pulses. Via a divider D and a pulse shaper I frequencies of 25, $12\frac{1}{2}$, $6\frac{1}{4}$ per second can be obtained. Hence it is possible in this way to get in the picture a variable number of situations per second which remain constant during a number of fields.

This arrangement was used to investigate the perceptibility of discontinuity as a function of contrast ratio and the velocity of the moving object. The result was rather surprising in the sense that perceptible distortion occurred much more quickly than was expected. For instance we found that a displacement speed of one picture width per two seconds, reproduced in 25 situations per second, showed a clearly perceptible and disturbing discontinuity at a contrast ratio of only 20. It must be concluded that contrasts of such a magnitude moving with such a speed seldom occur in films because they would not be acceptable to the audience. On the other hand we would add that as soon as the observer is acquainted to this effect, he perceives also much more discontinuities in normal film reproduction.

In the table the mean results from two observers are given as a function of contrast and speed. The room had no ambient illumination and the lowest from the two brightnesslevels was always 0.2 footlambert. The speed was adjusted twice for each value of contrast between background and bar, first so that discontinuity was just perceptible and then so that it was just acceptable.

For 25 situations per second and maximum contrast (160) the tolerable speed is one picture width in 12 seconds with respect to perceptibility and one picture width in 6 seconds with respect to acceptability. For $12\frac{1}{2}$ situations per second the contrast has to be decreased to $1/30$ th to make the same speeds acceptable. It can be concluded that, if contrast is kept constant and the rate of change is reduced from 25 to $12\frac{1}{2}$ situations per second,

- (a) the tolerable speed with respect to perceptibility is halved; and
- (b) the tolerable speed with respect to acceptability is reduced to $1/3$ rd.

The threshold of perceptibility for 25 situations per second is half the threshold for acceptability. For $12\frac{1}{2}$ situations per second the ratio is 1 : 3. For $6\frac{1}{4}$ situations per second only the lowest contrast (1,1) can make viewing conditions acceptable.

TABLE

Brightness in footlamberts		Critical speed in picturewidths per sec.					
backgr.	bar	25 sit. per sec.		12 ¹ / ₂ sit. per sec.		6 ¹ / ₄ sit. per sec.	
		perc.	acc.	perc.	acc.	perc.	acc.
0.20	32	<0.08	0.17				
0.20	16	0.14	0.22				
0.20	8.0	0.22	0.33		0.08		
0.20	4.0	0.25	0.50		0.10		
0.20	2.0	0.26	0.62		0.15		
0.20	1.0	0.27	0.66		0.17		
0.20	0.50	0.30	0.75	0.13	0.21		
0.20	0.25	0.33	0.75	0.17	0.25		≈0.08
0.25	0.20	0.35	0.75	0.18	0.27		≈0.08
0.50	0.20	0.33	0.66	0.17	0.25		
1.0	0.20	0.27	0.41	0.08	0.17		
2.0	0.20	0.25	0.37		0.12		
4.0	0.20	0.24	0.33		0.10		
8.0	0.20	0.20	0.26		0.08		
16	0.20	0.14	0.22		<0.08		
32	0.20	<0.08	0.12				

perc. = just perceptible

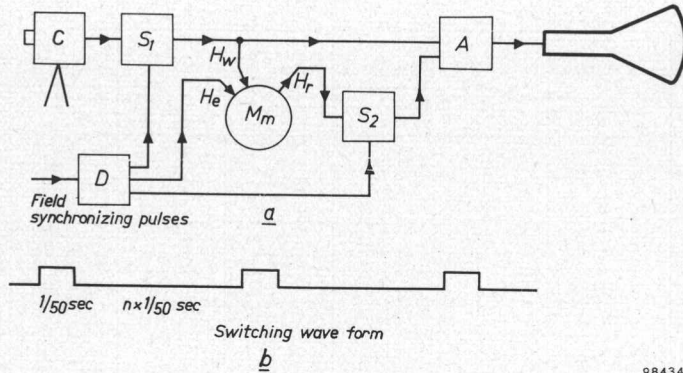
acc. = just acceptable

sit. per sec. = situations per second

Furthermore, it can be noted that for a small width of the bar the disturbing effects were somewhat greater than for a wide bar. Finally the experiment showed that a very disturbing impression occurred at about 6 situations per second. For 1 or 2 situations per second the steps are naturally much larger, but the effect is less disturbing because the observer has time to see the picture image as a series of separate stationary pictures.

Figure 16a shows the arrangement used for the investigation of normal pictures. From a signal source *C*, which could be a camera as well as a film scanner, the signal was fed to a switching device *S*₁. This switching device is operated by a voltage having a rectangular waveform, pulse widths correspond to a field period and the intervals between pulses to *n* field periods, *n* being an integer which can be chosen within wide limits. This switching voltage is derived

from the field synchronizing pulses of the signal source by means of a divider and is shown in fig. 16b. Owing to switching operation the video signal is only present at the output of S_1 during one out of every $n + 1$ fields.



98434

Fig. 16. Arrangement for reducing the number of situations per second in a television picture. C = camera, S_1 and S_2 are switching devices, D = frequency divider, Mm = magnetic memory device, H_w = writing head, H_r = reading head, H_e = erase head.

This output signal is fed to a magnetic memory device Mm of the kind designed by Backers and Wessels²²). This memory device consists of a rotating wheel with a magnetic coating on the periphery. A writing, an erasing and a reading head are placed opposite this. (H_w , H_e , H_r) The wheel makes 50 revolutions per second and is synchronized with the field frequency of the signal source. The output signal of H_r is coupled to a second switching stage S_2 operated complementary to S_1 . Consequently, during one field interval the signal is suppressed and during n intervals the video signal is passed through. A small phase difference with respect to S_1 is introduced in order to compensate for the distance between H_w and H_r . Besides S_1 and S_2 , the erase head is switched in accordance with the same pattern, and likewise with a suitable phase shift. Thus every $(n + 1)$ th field is recorded on the magnetic coating. This signal is picked up n times by H_r and then erased by H_e , after which a new field is recorded. The input signal fed to H_w and the output signal derived from H_r are added and then reproduced on a monitor. In this way a television picture can be represented showing a variable number of situations per second.

In the divider D seven binary counters are present, so that the lower limit is one new field per 128 fields. As is well known, it is possible by suitably adjusting feedback from the output to the individual counters to divide by all numbers between 1 and 128²³). A circuit diagram appears in fig. 17. Figure 17a shows an individual binary counter B with terminals I (input), O (output) and F (feedback). In fig. 17b a blockdiagram of the divider is shown. Each setting of the seven switches corresponds to some number between 1 and 128.

The signal sources which were used supplied signals in accordance with the 625-lines standard. Each field comprises $312\frac{1}{2}$ lines. This causes a certain irregularity because a phase shift of half a line will occur between two successive reading scans. A time base circuit is necessary in the monitor without flywheel synchronization, therefore, to keep the inherent distortion at the beginning of each field within acceptable limits. It has to be noted in addition that the definition of the reproduced picture in the vertical direction is only 312 lines, i.e. half the definition of a standard picture, but this does not affect the phenomena to be studied.

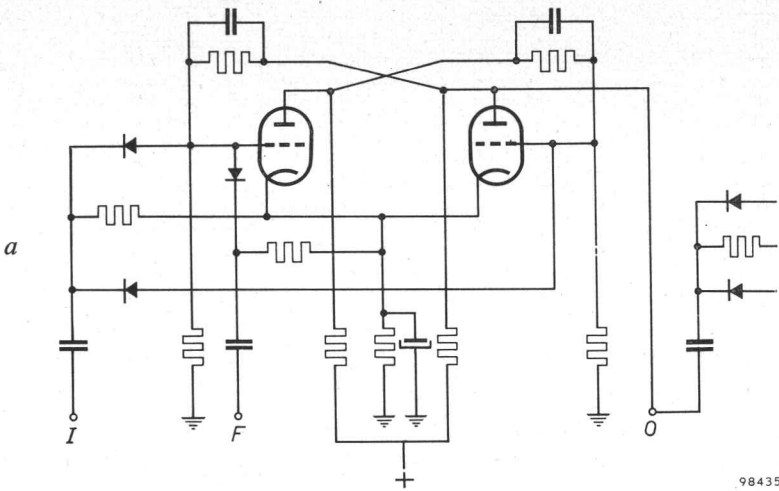


Fig. 17a. Diagram of a binary counter.

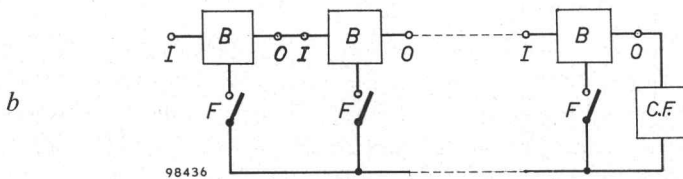


Fig. 17b. Block diagram of frequency divider with binary counters, I = input terminal, O = output terminal, F = feedback terminal, $C.F.$ = cathode follower.

After close examination of picture material of different kinds it was concluded that a critical region can be said to exist at 8 to 10 situations per second. Above this number most of the pictures observed were acceptable but below this limit the result is seldom acceptable. In these as in the first experiments, when the frequency of change is between 3 and 6 situations per second the disturbing effect is very strong. Below this rate the pure visual disturbance decreases, though the observer's impressions are naturally quite different from those of

normal television viewing. On the other hand, our experience indicated that very little can be expected from a long distance transmission — sometimes proposed — in which one or less situations per second are given. Certainly the program would have to be very especially adapted.

For the minimum figure in the above experiments a further decrease would be possible or, alternatively, better picture quality would be obtained, if a proper interpolation between successive picture situations could be established. In the experiment described the interpolation can be said to have been carried out in such a way that in the interval in which the real picture situation is unknown the picture is kept identical to the last known data. This in fact is a very simple interpolation, which can be said to give a first approximation of the real picture. A continuous fade-over from each situation to the following one is a second approximation. The question arises as to whether this would involve a real improvement. It is true that abrupt changes would be eliminated but, in addition, moving parts would suffer a certain loss of sharpness. Nevertheless an experiment on these lines will be of interest. Further approximations may be imagined approaching more and more the interpolation which can be carried out undoubtedly very well by the human observer if he had the opportunity to study closely successive picture situations. The achievement of this interpolation by technical means seems hardly possible other than by complete electronic computers. Yet some investigators make an attempt to use this principle for bandwidth reduction ²⁴).

When 10 picture situations per second are employed the normal number of 50 is divided by 5. This does not apply however, to the bandwidth reduction, because in fact the line-interlace principle in standard television constitutes in itself a halving of the required bandwidth. Therefore only a gain given by a factor $2\frac{1}{2}$ results.

It will be clear that in our experiment no bandwidth compression was achieved. To reduce the bandwidth in practice it is necessary to make 1/10th of a second available for the transmission of one complete frame. It is scarcely necessary to point out that if a transmission system of this kind were put into practice, flicker free reproduction at the receiver involves considerable technical difficulties. We shall go further into this subject in III.2.

II.3.2.2. A decrease of the information per field

Apart from a decrease of the number of fields per second the number of complete frames can also be reduced by using more than two fields to transmit information about a complete set of picture elements. This could be done by extending the principle of line-interlace. However better results may be expected when a process similar to that used in line-interlace is also employed in horizontal direction for the picture elements.

Suppose the picture is split in picture elements (dots) instead of lines. A single scan deals with only half the number of dots in a line. The remaining dots in that line are dealt with in the next scan. When both dot series are interlaced like the two line patterns of successive fields, picture elements are transmitted at half the normal rate, and therefore only half the bandwidth is required. (The special conditions necessary to allow bandwidth to be reduced thus are considered in III.3.1.) Of course, in this system all picture elements are transmitted in four fields instead of two and the number of data per picture element per second is also half the normal one. Yet a definite difference exists between this and a system in which half the normal field frequency is applied. In the latter case 25 complete frames and 50 fields are transmitted every two seconds. In the dot-interlace system 25 complete frames are also transmitted per two seconds, but in the form of 100 fields with line- and dot-interlace. In the first system the normal sharpness is maintained in each field but the observer has to accept the representation of 25 fields per second as a picture continuously present. In the second case 50 fields per second are still present as in normal television. Hence large areas without detail will show no flicker. However, the observer is now asked to accept the appearance of detail split between two partial reproductions.

It can be easily shown experimentally that the dot-interlace system is preferable. However, in this system too, the visual impression differs considerably from that given by a standard television image. A detailed description of dot-interlace picture quality will be deferred until the system as a whole is described in III.3.1, because it calls for a clear insight into the transmission process. One particular aspect can however be described here.

As stated in I.2, the picture quality perceived depends on the observer's eye movements, and therefore these always have to be taken into consideration. On the other hand besides the picture content, there exists an influence of the reproduction process on the eye movements as well. Line crawl in normal television may be mentioned as such an effect. In dot-interlace the stroboscopic phenomena caused by successive dot patterns, each positioned somewhat differently from the previous one, give rise to dot crawl which is much more striking. When the observer fixes his eyes on a particular spot the dot structure may be hardly visible, but in normal viewing moving dot patterns seem to arise, diverting the observers attention from the picture content in an irritating way. As the principle of dot-interlace only prescribes the relative positions of the dot series in two successive scans of the same line, it is possible to establish relative positions of dot series in adjacent lines such that the visibility of stroboscopic effects is reduced to a minimum. For this purpose some experiments were carried out ²⁵).

An interfering sine wave with a frequency equal to half the number of picture elements transmitted per second in the 625-lines television standard was super-

imposed on a video signal. The sine wave was of sufficient amplitude to cause dot series on each line in the reproduced picture. The phase of the sine wave could be adjusted in several ways but the phase difference for two successive scans of the same line was always maintained at 180° . The phase relationship for successive lines of one field and of adjacent lines in the picture could be adjusted to 0° , 90° or 180° . This was achieved by means of an oscillator O (see fig. 18) which was started at the beginning of every line and stopped at the end. In addition, the oscillator voltage was fed to two commutators C_l and C_f operated at line and field frequency respectively and connected in series. They established a line and/or field alternating phase shift (in P_1 and P_2) of 90° or 180° . The output signal which had passed the two commutators showed for each line the desired phase.

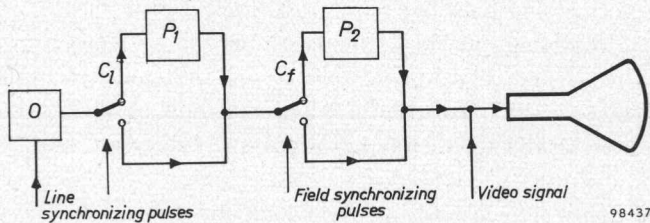


Fig. 18. Arrangement for obtaining a special relationship between the phase of an interfering sine wave and the scanning frequencies. O = oscillator, C_l = commutator operated in line frequency, C_f = commutator operated in field frequency, P_1 and P_2 are phase shifters.

The dot patterns which were examined are represented in fig. 19. Here the relative positions of bright areas corresponding to maxima of the sine waves have been denoted as follows:

- in the first field by A (odd lines);
- in the second field by B (even lines);
- in the third field by C (odd lines);
- in the fourth field by D (even lines).

The principle of dot-interlace requires that on a certain television line the dots A are midway between the dots C and the dots B midway between the dots D , but for the rest the relative positions can be freely chosen. Now if dots A , B , C , D are located on a straight line a stroboscopic effect may easily arise, because the four impressions A , B , C and D occurring at intervals of $1/50$ th of a second may easily be interpreted as one moving dot pattern. This happens in fig. 19a and b. In fig. 19a it is indicated by arrows. In fig. 19c and d it is impossible to locate A , B , C and D on a straight line. The stroboscopic effect just referred to will not occur. However, another stroboscopic effect may accompany this configuration, namely a moving pattern which can be said to be of the second order. If the eye moves in such a way that the area of attention shifts half the

dot distance $A - A$ every $1/25$ th of a second, in horizontal direction, the impression of dot pattern A will coincide with that of dot pattern C , the impressions of B and D likewise coinciding. (For some dot patterns — fig. 19a and b — this can also happen in vertical direction.) In this case two moving patterns $A = C$ and $B = D$ are observed. The pattern perceived in this way is represented in fig. 20 for the dot pattern of fig. 19d. As the fields coincide in two's, the brightness of the second-order moving patterns will be less than that of the first-order patterns present in the configurations of fig. 19a and b.

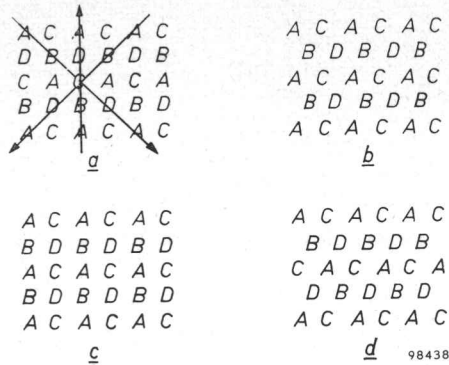


Fig. 19. Dot patterns of interfering sine waves in a television picture. First field: A , second field: B , third field: C , fourth field: D .

An analysis with respect to stroboscopic effect shows that the second-order effects in fig. 19d appear in both the vertical and horizontal directions and are equally perceptible in all four directions. Further the moving patterns are finer than those of the patterns of fig. 19c. Therefore fig. 19d has the most preferable configuration with respect to visual disturbance. Whether the patterns can be produced in a simple way is another matter. Figure 19a is the best from the viewpoint of simplicity. It can be easily shown that this pattern is obtained without further precautions if the sine wave frequency equals an odd multiple of half the line frequency. We shall return to this question in III.3.5, where experiments on systems are described.



Fig. 20. Moving pattern which may be observed in fig. 19d.

The disturbing effects just discussed are not merely of interest in dot-interlace systems. They occur in a similar way when subcarriers are employed to increase the efficiency of a transmission system. These subcarriers are modulated with a television signal and transmitted within the frequency band of another television signal. In order to reduce the visual effect of the subcarriers in the image obtained from the second signal, the extraneous sine wave is given opposite polarities in alternate frames. The same considerations apply to phase relations in adjacent lines in the picture as given above for dot-interlace. The application of subcarriers is described in more detail in sub-section III.3.2.

By the employment of dot-interlace or subcarriers, as described above, the frame frequency can be said to be halved because the complete undisturbed picture information is given by four fields instead of two. The question may arise as to which degree it is possible to extend these principles thus that a frame is transmitted by more than four fields. Proposals have been made in this line ²⁶). However, experiments which we carried out have shown that it is quite impossible to get an acceptable picture when the visual integration has to extend over more than four fields.

II.3.3. *The differential sensitivity of the eye*

Not only the resolving power and the speed of visual perception are limited, but also the ability to perceive differences in brightness levels. Rapid transitions as well as slow variations as a function of place can be only perceived beyond a certain threshold. The ability to see differences in brightness will be referred to here as "differential sensitivity". We shall be giving a rather wide interpretation to this term, using it to cover all effects which are related to the perceptibility of and to the disturbance caused by local errors in brightness reproduction. Strictly speaking, then also properties of the eye other than the ability to recognize small differences of contrast are involved.

Differential sensitivity is not so closely related to bandwidth as resolving power and persistence of vision are. Only after a special encoding process can the number of steps in contrast be correlated to the channel bandwidth needed. On the other hand, the differential sensitivity is directly related to the signal-to-noise ratio which can be tolerated, and it determines the "metrical information" of a picture element per scan. It will be clear that signal-to-noise ratio is as important as bandwidth, and if the signal-to-noise ratio can be improved it is not necessary to transform this in a bandwidth gain in order to make its advantages evident.

The limitations on differential sensitivity make it possible to accept a certain amount of error in picture reproduction. Now, it may be asked if differential sensitivity has any special aspects which enable transmission requirements with respect to bandwidth or signal-to-noise ratio to be further reduced. Three aspects can be noted here.

- (1) The disturbing effect of noise decreases with increasing frequency ^{27,28,29}).
- 2) The disturbing effect of noise of certain amplitude decreases with increasing brightness level ³⁰).
- (3) The disturbing effect of noise decreases with an increasing amount of picture detail.

The first two aspects are well known and do not call for further discussion. At special points the television practice is matched to these effects. For instance, the camera-tube output signal is usually amplified in such a way that noise shows an increasing energy density with increasing frequency. In addition, the non-linearity of the picture tube characteristic causes a relative decrease and increase in transmission noise in dark and bright areas respectively. The third effect — less disturbing impression if more picture detail is present — is not often mentioned, probably because its technical consequences are less obvious. The investigations and proposals of Schönfelder, Kretzmer, Schreiber and Graham can be said to be more or less based on this effect. We shall give here a brief description of their publications.

Schönfelder investigated "frequency dependent gamma correction" ³¹). This principle involves a type of gamma correction — as required for signals derived from a flying-spot scanner — which is frequency-dependent in the sense that only the lower frequencies are corrected. In consequence, the high frequency components of flying-spot scanner noise are not amplified to such a large extent in dark areas as normally occurs in gamma correction. The visibility of noise is thereby reduced. On the other hand, of course, a mismatch will be present between the lower and higher frequencies of the picture signal, as they are treated in a different way. So errors will occur in picture detail for better signal-to-noise conditions in large dark areas.

Kretzmer has carried out experiments on a system in which a quantized television signal was employed and in which coarse quantization was used for the high frequencies of the video signal ³²). He was able to show that the picture was still acceptable when frequencies above 0.5 Mc/s were quantized at only 5 levels. On the other hand 8 levels for the whole frequency range was not acceptable. Here too a discrepancy is found between the differential sensitivity of large areas and of picture detail. The saving provided by this system can be evaluated — at least from a certain point of view — if it is assumed that binary-pulse coding is used for the transmission signal. The required bandwidth is then 60% of the bandwidth required for binary-pulse coding transmission with equal quantization for all frequencies.

Schreiber uses the principle of Kretzmer for a system in which the correlation between adjacent picture elements is also exploited for reducing channel capacity ³³). This correlation is only measured for the high frequency content of the picture signal and only for that part of the signal the number of picture elements transmitted is limited. Calculations are given for a transmission signal

encoded in binary pulses. A quarter of the normal channel capacity is needed, no technical proposals are given for the realization of standard channel bandwidth compression.

The system of Graham can be said to be based on a more general interpretation of the relative small differential sensitivity in picture detail ³⁴). He describes this perceptual effect as a small differential sensitivity for areas with unexpected picture content (where the observer is surprised) and a large differential sensitivity for areas where picture content can be easily predicted. Accordingly, in his system the difference is transmitted between the real picture signal and a signal such as can be predicted from the foregoing. This difference signal is quantized in small steps for small values and in large steps for large values of the difference signal, the amplitude of which can be said to be a measure of the "surprise" of the observer. The experiments of Graham show that 8 levels for the difference signal are sufficient for an acceptable image. This means that 3 bits per picture element are sufficient owing to the perceptual characteristic referred to above. However, as a difference signal is transmitted and prediction is involved, an exceedingly reliable digital transmission is required.

Apart from the ideas of various investigators as just described a more straightforward attempt is conceivable to eliminate the difference in visual interference by noise in large areas and in picture detail. The reproduction of high frequencies could be suppressed as long as high video signal frequencies can be assumed to be absent. In a region without any detail the output signal of a high-pass filter will be on a low level. This may be used to measure picture detail. Of course, noise is also present in this output signal. It may be possible, by continuous measurement of the r.m.s. value for a time interval chosen, to distinguish between noise and signal plus noise. As long as no video signal is present the reproduction of the high pass filter signal has to be omitted. Filter characteristic as well as the r.m.s. measurement has to be dimensioned in such a way that a delayed action of the electronic switch is avoided and the erroneous blocking of the high frequency path is kept within acceptable bounds.

It will be clear that this technique — like all the measures that have been mentioned in this section — only produces a fairly small improvement in signal-to-noise conditions. They do not offer any remedy for extremely bad signal-to-noise conditions. The question arises as to which method has to be employed for transmission if the signal-to-noise ratio is very low and the normal bandwidth is available. Though an extensive treatment is outside the scope of this study some remarks will be given.

When the signal-to-noise ratio is very low it is necessary to employ modulation systems that are less sensitive to noise, such as frequency modulation and pulse modulation. As a large relative bandwidth is needed for these systems, the video-signal bandwidth has to be reduced. Hence a part of the information

will be omitted in order to reduce noise interference and to obtain a better overall picture quality. This also holds if a very coarse quantization of brightness values is used for the elimination of noise. It is well known however that picture quality rapidly decreases when the number of levels falls below 32³⁵). For a very high noise level the resulting loss of picture quality is certainly too excessive.

However it may be possible to make a better use of the quantization. For instance it seems better, in the extreme case when two levels are employed, to transmit and reproduce only picture contours instead of black and white areas. Since in normal life a great deal of visual information is conveyed by outlines, it may be assumed that usually a more acceptable result will be obtained with contour reproduction than from the reproduction of all brightness levels below the centre level as black and all above this level as white. The location of the contours can be preferably chosen at the points of inflexion in the waveform of the normal video signal. A selection has to be made so that only those points will be recognized as contours where a certain minimum slope is exceeded.

It is of interest to examine the improvement in signal-to-noise ratio resulting from two-level transmission of this kind. It may be assumed that the picture is acceptable as long as the noise level is above the threshold (half the signal range) in less than 5% of the picture elements. This figure of 5% is based on our experiments in which we investigated the visual disturbance arising from the introduction of interfering pulses with random distribution over the picture area, and a width of one picture element. In noise with a Gaussian distribution twice the r.m.s. value is exceeded 5% of the time. Hence the ratio between the signal range and the r.m.s. value of the noise has to be at least 4, which is 12 dB. In normal television 30 dB is just acceptable, and it follows from this reasoning that a gain of 18 dB is obtained.

II.3.4. *Perception of colour*

Up to now we have only been concerned with the transmission of picture brightness. In this sub-section we are going to investigate redundancy in the transmission of colour information.

As is well known, at least three independent quantities are necessary in order to determine the brightness and colour of a certain light impression³⁶). Hence if the number of picture elements in colour television were made the same as in monochrome television, three television signals would be required, because three data per picture element have to be transmitted instead of one. However, it is not a priori certain that the accuracy required for the reproduction of colour as a function of place in the television picture is the same as for brightness; the choice of the number of picture elements in monochrome television has been based on perception of brightness only. In order to determine the required accuracy for colour, and so the required bandwidth, additional data about

perception are needed. Some of this information has long been available. During the development of colour television it was rediscovered and supplemented.

In general, it may be stated that if normal bandwidth is employed for one signal suitably chosen from amongst the three signals needed for colour-television transmission, one or both of the other signals can be transmitted within a much narrower bandwidth. Quantatively it depends to a large extent on how the signals are made up. Various possibilities exist: for instance the signals may refer to the green, red and blue light component, or to luminance, hue and saturation or to combinations of these quantities. In most proposed systems signals are transmitted which are proportional to the green, red and blue light content or to linear additions of these primary signals, as they are called. The reason is, that it is possible to reproduce almost any coloured light impression by a proper addition of — suitably chosen — green, red and blue light.

The required definition for the individual colours red, green and blue in colour picture has been investigated by Baldwin³⁷⁾. A colour picture was observed which was obtained by the optical addition of three primary pictures in green, red and blue. Definition requirements were ascertained by defocussing the primary pictures. It was found that about the same definition is required for the green picture as for a monochrome image whereas for the red picture it is permissible to have a lower, and the blue image a much lower definition. This holds when the green, red and blue pictures are combined to a composite colour picture. If the green, red and blue picture are observed separately, the differences in definition required are not very large.

Previously Bedford had noted that the resolving power of the eye is less for colour detail than for brightness detail³⁸⁾. He first proposed a transmission system in which the three primary signals were transmitted separately up to a certain frequency, being transmitted as one composite signal above this frequency. This means that the receiver will not show picture detail in the correct colour. However, if the composite high-frequency signal is made to convey luminance information, picture detail will be reproduced with the correct luminance values. The signal in question and in general a signal which is proportional to the luminance is obtained by a suitable addition of the primary signals. For the signal in question of course only the high frequencies are added. (Known as mixed highs.)

As the analysis of the transmission problems progressed it became clear that not merely the transmission of high frequencies of the luminance signal is preferable, but that the transmission of the whole luminance signal as the main signal is advisable for several reasons^{39,40)}. Apart from this main signal two colour signals have to be transmitted, and these may have a limited bandwidth. Henceforward investigation on colour perception concentrated on ascertaining the minimum requirements for the additional colour information. These requirements again depend on the composition of the colour signals.

The results of the investigations into colour sensitivity carried out by Willmer and Wright are very important in this respect ⁴¹). They found that for small areas a certain colour impression can be obtained by the suitable addition of only two colours. Whereas for large areas the addition of green, red and blue is needed, for small areas almost all colour impressions can be produced by a proper addition of orange and cyan. From these findings, and from the principles set out above, we may draw the conclusion that:

- (a) three signals are needed for the transmission of large areas in the colour picture,
- (b) two signals are needed for colour picture detail,
- (c) one signal is necessary for the very fine detail.

The transmission system now in use in U.S.A. is based on this ^{42,43}). Three signals are used, with three different bandwidths. The large-bandwidth signal is the luminance signal. Its bandwidth is the same as that of a normal black and white signal (4 Mc/s). The second signal is called the “I signal”. This signal refers to the position of the actual colour in the colour triangle with respect to the complementary colours orange and cyan. The bandwidth is about 1/3 of normal bandwidth (1.5 Mc/s). The third signal — called the “Q signal” — refers to position in the colour triangle with respect to the complementary colours green and magenta. The bandwidth is about 0.5 Mc/s. Hence up to 0.5 Mc/s three independent signals are available, between 0.5 Mc/s and 1.5 Mc/s two signals are present, and beyond 1.5 Mc/s only the luminance signal is available for transmission.

Davidse pointed out that apart from the resolving power of the eye for colour detail, the statistical distribution of colour differences plays an important role in determining the bandwidths required for the three signals ⁴⁴). He showed by statistical measurements that, for normal pictures, transients in the *Q* signal — corresponding to transients in colour from green to magenta — do in fact occur less frequently than transients in the *I* signal corresponding to transients in colour from cyan to orange. In a test signal involving similar transient conditions for *Q* and *I*, a larger bandwidth is needed for the *Q* signal. This aspect of redundancy in colour transmission in fact belongs in section II.2.2.1.

Apart from the resolving power, the influence of noise on colour picture quality is naturally of very great interest. This influence depends on the eye's differential sensitivity to colour. This sensitivity has been extensively investigated by Mac Adam ⁴⁵). It has been determined for a large number of positions in the colour triangle. With the aid of these data the noise characteristics of a colour television transmission system can be evaluated ⁴⁶). Of course a situation is preferred in which large differences in error sensitivity to various colours are avoided.

We want to finish this section with an additional remark about statistical redundancy in colour transmission. For natural pictures the picture content

consists of limited number of objects each having a nearly constant hue and saturation. In the colour-television transmission systems in use, therefore, a large amount of redundancy remains, notwithstanding the ingenious solution of the bandwidth problem. Once the luminance signal is given, in fact, only two data are needed per object. Also this kind of redundancy can only be exploited by adopting more complicated systems of transmission.

II.4. Psychological limitations

At the end of our analysis of redundant aspects in picture transmission some remarks will be made about what we have called psychological limitations.

Apart from the characteristics of the eye, there is a certain limitation on apprehension by an observer in the sense that he needs a certain time to recognize or to get acquainted with the objects in the picture. Sziklai has investigated how much time an observer needs to recognize well known objects in a television reproduction ⁴⁷⁾. He found a time interval of about a quarter of a second per object. As a total number of about 1000 was assumed for such simple well known objects, the capacity of the human visual channel was calculated to be 40 bits per sec.

As stated in the general introduction, our opinion is that one has to be very cautious in interpreting these results. Certainly they give an idea of the relatively small rate at which factual data are assimilated. However, these results are isolated from conditions which have to be fulfilled if the assimilation of pictorial material is to be a pleasurable experience. One condition was noted in the introduction, namely that each observer should be free to concentrate his attention on any part of the picture. Each observer will scan the picture in a different way, and will get different impressions though the factual data are identical for all observers. In terms of information theory, it may be said that the channel capacity of visual perception is very small but the amount of information supplied per second in transmitting images has to be related to a large number of observers.

Apart from this circumstance of course certain aesthetic aspects play an important role. As stated in II.3.3, the contours in a picture carry a large part of the factual information; this is illustrated by the fact that any drawing is an adequate means for getting a clear impression of a configuration. However, shaded areas contribute to a large extent to "picture quality". This is a simple example of certain factors which are important for the overall sensations but difficult to express in bits/second.

In addition, the important phenomenon of large error sensitivity for well known configurations has to be pointed out. The observer tends to pay attention to that part of the picture which shows deviation from well known and therefore predictable geometry. However this part of the picture may be other than that which is essential for watching the picture program. Therefore the

disturbance to picture parts which are unimportant to the plot, and which might be said to be redundant, can be irritating due to the diversion of attention. In this connection the paradoxical statement might be made that a large part of the television picture is indeed redundant for psychological reasons, but only as long as it is present.

Apart from the considerations given above, which make it difficult to determine and avoid psychological redundancy, there is the fact that the observer at the receiver has a considerable store of knowledge about objects, situations etc. with a high degree of probability, so that a further element of redundancy arises. It will be clear that within the scope of this work it is not useful to go further into this subject though it is very interesting from a purely perceptual point of view. Foregoing remarks have been given mainly in order to point out once again the danger of drawing conclusion from too straightforward a comparison of the actual television transmission and data about visual perception derived from information theory.

REFERENCES

- 1) S. Deutsch, Trans. Inst. Radio Engrs **BTR-2**, 69-82, 1956.
- 2) H. Schober, Optik **13**, 350-364, 1956.
- 3) C. Harrison, Bell Syst. tech. J. **31**, 764-783, 1952.
- 4) E. Kretzmer, Bell Syst. tech. J. **31**, 751-763, 1952.
- 5) W. Schreiber, Probability distributions of television signals, Thesis, Harvard University, 1952.
- 6) W. Schreiber, Trans. Inst. Radio Engrs **I.T. 2**, no. 3, 94-105, 1956.
- 7) N. Deriugin, Telecommunications, **7**, 1-12, 1957.
- 8) Y. Galitskaya a.o., Radio Engng, **12**, no. 3, 73-77, 1957.
- 9) K. Powers and H. Staras, Commun. Electronics **32**, 492-496, 1957.
- 10) E. Cherry and G. Gouriet, Proc. Instn. electr. Engrs **100 III**, 9-18, 1953.
- 11) G. Gouriet, J. Televis. Soc. **8**, 103-108, 1956.
- 12) G. Gouriet, Proc. Instn. electr. Engrs **104 B**, 265-272, 1957.
- 13) D. Novik, Sovjet Physics — Techn. Physics **1**, 886-897, 1957.
- 14) C. Shannon, Bell Syst. tech. J. **27**, 379-423, 623-656, 1948.
- 15) F. v. d. Poel and J. Valetton, Philips tech. Rev. **15**, 221-232, 1953-54.
- 16) G. Gouriet, Electron. Engng **24**, 308-311, 1952.
- 17) F. Schröter, Arch. elektr. Übertr. **7**, 63-70, 1953.
- 18) H. Hartridge, Recent advances in physiology of vision, p. 19, Churchill, London, 1950.
- 19) W. Duke, Textbook of ophthalmology, p. 589, Kimpton, London, 1946.
- 20) H. de Lange, Attenuation characteristics and phase-shift characteristics of the human fovea-cortex systems in relation to flicker-fusion phenomena, Thesis, Delft, 1957.
- 21) J. Haantjes and F. de Vrijer, Wireless Engr **28**, 40-42, 1951.
- 22) F. Backers and J. Wessels, to be published in Philips tech. Rev.
- 23) B. Chance a.o., Waveforms, Sec. 16, Mc. Graw-Hill, New York, 1949.
- 24) D. Gabor, Television compression by contour interpolation, Lecture, Course on Inf. Theory, London, 1958.
- 25) K. Teer and J. Valetton, Unpublished report, Philips Research Lab., 1952.
- 26) F. Thompson and P. Toulon, Conv. Rec. Inst. Radio Engrs **2**, part 7, 153-164, 1955.
- 27) J. James, Proc. Instn. electr. Engrs **99-III A**, 796-803, 1952.
- 28) J. Barstow a.o., Commun. Electronics **72**, 735-741, 1953.
- 29) M. Gilbert, B.B.C. Engng Divn, Monograph 3, part 2, 1955.

- ³⁰⁾ R. Theile a.o., Arch. elektr. Übertr. **10**, 98-104, 1956.
- ³¹⁾ H. Schönfelder, Arch. elektr. Übertr. **10**, 512-534, 1956.
- ³²⁾ E. Kretzmer, Conv. Rec. Inst. Radio Engrs **3**, part 4, 140-147, 1956.
- ³³⁾ W. Schreiber a.o., Conv. Rec. Inst. Radio Engrs **6**, part 4, 88-99, 1958.
- ³⁴⁾ R. Graham, Wescon Conv. Rec. Inst. Radio Engrs **2**, part 4, 147-157, 1958.
- ³⁵⁾ W. Goodall, Bell Syst. tech. J., **30**, 33-49, 1951.
- ³⁶⁾ P. Bouma, Physical aspects of colour, Philips tech. Library, Eindhoven, Netherlands, 1947.
- ³⁷⁾ M. Baldwin, Proc. Inst. Radio Engrs **39**, 1173-1176, 1951.
- ³⁸⁾ A. Bedford, Proc. Inst. Radio Engrs **38**, 1003-1009, 1950.
- ³⁹⁾ B. Loughlin, Proc. Inst. Radio Engrs **39**, 1264-1273, 1951.
- ⁴⁰⁾ W. Bailey, Proc. Inst. Radio Engrs **42**, 60-66, 1954.
- ⁴¹⁾ E. Willmer and W. Wright, Nature **156**, 119-121, 1945.
- ⁴²⁾ Ch. Hirsch, Advanc. Electron. **5**, 316-365, 1953.
- ⁴³⁾ G. Brown, Proc. Inst. Radio Engrs **42**, 58-59, 1954.
- ⁴⁴⁾ J. Davidse, to be published in Electronic and Radio Engineering.
- ⁴⁵⁾ D. Mc. Adam, J. opt. Soc. Amer. **32**, 247-274, 1942.
- ⁴⁶⁾ F. de Vrijer, Acta Electronica **2**, 103-109, 1958.
- ⁴⁷⁾ G. Sziklai, Trans. Inst. Radio Engrs **I.T. 2**, 125-128, 1956.

III. TRANSMISSION SYSTEMS WITH BANDWIDTH COMPRESSION

III.1. Introduction

In this part investigations into systems with bandwidth compression will be dealt with. In the previous part the statistical and perceptual aspects of picture transmission were accentuated; here the properties, characteristics and typical equipment of transmission systems will be considered. The considerations will be restricted to those transmission systems on which we carried out practical experiments. Therefore only a few principles of the great many of principles which can be proposed and have been proposed for television transmission within a narrow frequency band will be discussed. The systems which will be described are all based on a reduction of the number of frames transmitted per second which is in accordance with the results of the analysis given in the previous part. The results given in part II in fact lead to the conclusion that the most promising possibilities for bandwidth compression will be found in a reduction of the number of frames transmitted per second, where frame is used to mean one complete cycle of the scanning process. This reduction can be effected in two ways.

In the first place the number of fields per second, that is, the number of cycles in vertical deflection per second, can be decreased. This gives rise to a flicker problem which has to be solved.

In the second place the information transmitted per field can be reduced so that more than two fields are required for the transmission of one complete frame. For instance this is the case where the principle of dot-interlace, as described in II.3.2.2, is employed. In this case no increase of flicker will be perceived in large areas of the picture, but disturbing effects will occur in small regions which have to be reduced to a sufficiently low level. Moreover in most cases extra synchronization arrangements have to be made between transmitter and receiver.

We shall first give attention to the former method, a decrease in the number of fields per second.

III.2. A decrease in the number of fields per second

The design of a system with a reduced field frequency involves serious problems. At the receiver special precautions have to be taken in order to avoid flicker which will be present if the scanning rate is below a limit of about 50 fields per second. Either the receiver must have special arrangements for light

excitation that maintains the brightness level throughout the frame period with a sharp cut-off at the end, or a special device is necessary in which a signal with the normal field frequency is derived from the incoming signal with a slow scanning rate.

Both possibilities are far from practical. In regard to the former, one practical proposition exists in the form of electro-luminescent display panels ¹⁾. The alternative would be to use a memory device in which the information is stored in a slow scanning rate and is recovered by scanning at a normal speed. To put it more precisely a double memory device would be necessary, for the information in any given frame could not be recovered until the whole of the frame had been stored. This would apply where the change-over from one frame period to the next had to be instantaneous. The entire content of the frame would then have to be available at any given instant, and this would only be possible with two memory tubes in tandem. See fig. 1.

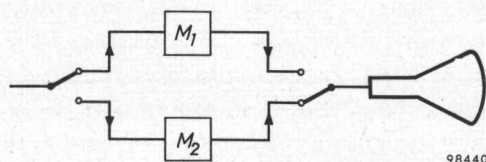


Fig. 1. Video-section of a receiver of a slow-scanning transmission system. The received signal is written alternately in M_1 and M_2 (slow-scanning rate). The display signal is read alternately from M_2 and M_1 (normal scanning rate).

The signal content of a certain frame period of the incoming signal is written into the memory device M_1 . In the same time interval the content of the previous frame period — present in M_2 — is read with normal frame frequency and fed to the display device. At the end of the period, M_1 is switched in the reading position and connected to the display whereas M_2 is switched in the writing position and connected to the input. As the frame frequency in the transmission is reduced each writing scan will correspond to several reading scans. Hence repeated reading of the memory devices must be possible.

Special difficulties may also arise on the camera side ²⁾. The sharpness of moving objects in the picture may suffer because the integration interval of the camera tube may become too long. In this case the exposure of the photo-sensitive target has to be instantaneous and the camera tube must also be able to deliver signal in the unexposed intervals. Most types of camera tube are in fact able to do this when suitable measures are taken. On the other hand a decrease of the signal to noise ratio in the camera output — under constant lighting conditions — will be inevitable when the relative integration interval is decreased. At the receiver, however, the problems are most difficult.

The memory device which is needed at the receiver side may be based on electrostatic or magnetic recording ^{3,4)}. What has been realized in this field

up to now is far from being perfect either with respect to quality or with respect to reliability and simplicity.

As the practical application of the systems under discussion here is hindered by the complexity of memory devices it seems advisable to start by studying those transmission systems which involve the simplest storage problem. Once the simple storage problem is solved, so that a simple but perfect memory device is available, an attempt can be made to obtain more complex memory functions by combining several of these separate "memory units" as circuit elements in a suitable arrangement with the application of normal network principles, such as connection in series, connection in parallel, feedback and so on. Possibly this is a better approach for the solution of the memory problem, (which is one of the essential difficulties in bandwidth compression and which occurs in far more systems than merely those systems based on a decrease of field frequency) than the construction of one complicated memory tube or similar device. According to this reasoning we looked out for transmission conditions which make the receiver storage problem as simple as possible.

The problem at the receiver becomes somewhat less complicated if multiplex transmission is being contemplated, and especially if two television signals are to be transmitted in one normal channel. In this case the normal frame period can be maintained by halving the number of frames per second and by transmitting frames corresponding to the two signals alternately. All that is required at the receiver is equipment which makes it possible to reproduce the content of the preceding frame during the frame interval which does not carry the desired signal. The memory device then can be far more simple because the speed for writing and reading can be the same. In principle a television camera tube can be used for the purpose ⁵⁾. Obviously the most simple type of camera tube will be chosen and that suggests the vidicon type ⁶⁾.

The following reasoning will make clear how a camera tube can be made to perform the memory function. It holds for the vidicon type; for the image iconoscope, image orthicon and so on similar principles apply.

During the scanning of the photosensitive target the potential of each target element is raised to cathode potential or to a potential approximating the cathode potential. Hence if the cathode potential varies during the scan the corresponding variations of target potential will appear over the area scanned. When a television signal is fed to the cathode a corresponding image will be present as a charge pattern on the target after one complete scan. During the succeeding scan a signal current component will be found corresponding to this charge pattern, and hence to the picture content of the preceding frame. Apart from this, at any given instant a current component will be present which is directly related to the instantaneous value of cathode potential. Therefore the output current of the camera tube will show a close relationship to the difference between the picture content of two successive frames.

In principle by proper matrixing of in- and output signal a signal can be obtained which shows a delay of one frame period. Now, if the input signal only carries information about the desired signal during the odd frame periods, the information missing during the even frame periods can be replaced by the delayed signal of the odd frame periods. The diagram in fig. 2 represents a system for the transmission of two television signals in the normal channel bandwidth, based on this principle.

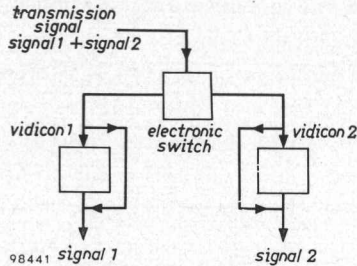


Fig. 2. Video-section of a receiver of a duplex transmission system with normal scanning frequencies. Frames corresponding to the two signals are transmitted alternately. In the reproduction of each signal the missing frames are replaced by the delayed preceding frames. This is achieved by means of an electronic switch and a memory device (vidicon).

Frame periods of two different video signals occur alternately at the input. For instance the signal 1 is present in the odd periods and the signal 2 in the even periods. By means of an electronic switch the input signal is suppressed during the frame intervals carrying the undesired signal: for the reproduction of signal 1 (left hand branch) the even periods are suppressed, for signal 2 (the right hand branch) the odd periods. After this process the vidicon in each branch establishes a difference signal of two successive frame intervals. Now in this difference signal the desired signal is present in all frame intervals but with alternating polarity in successive intervals. By proper matrixing of the vidicon input- and output signal a suitable display signal can be achieved with constant polarity in which the signal content of the odd periods is repeated in the even periods (for the signal 1), or even periods are repeated during odd periods (for the signal 2).

The system just described was the subject of an investigation. It will be clear that in this system too the storage process is essential and the rest of the transmission chain involves no particular technical difficulties.

Now first a theoretical analysis of the essential system device — the vidicon used as a memory tube — will be given in 2.1. In 2.2 follows a description of practical experiments. Here especially the use in duplex transmission — according to fig. 2 — is considered.

III.2.1. Theoretical analysis of the vidicon used as a memory device

Figure 3 is a diagram showing the layout of the vidicon. An electron gun G emits an electron beam B which scans the photosensitive target T . This is affixed to a conductive transparent plate P connected by the resistor R_s to a voltage source of value V_p . The target is supposed to be uniformly illuminated. The cathode of the electron gun is connected to the signal source V_k which represents the input signal. The electron beam is focussed and deflected by suitable magnetic fields so that the target is scanned by the electron beam according to the well known television line pattern. The signal current i is that proportion of the beam current which is taken up by the target. The output signal V_u appears at the top end of R_s :

$$V_u = V_p - iR_s. \quad (1)$$

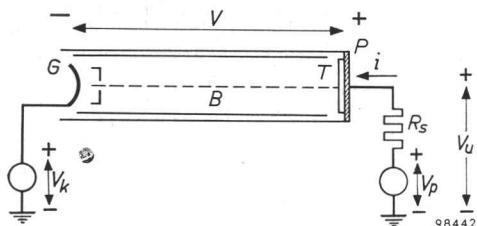


Fig. 3. Diagram of the vidicon used as a memory device. G = electron gun, B = electron beam, T = target, P = plate, R_s = signal resistor, V_k = input voltage, i = signal current, V_p = plate voltage, V_u = output voltage.

Now this arrangement of fig. 3 can be considered as a four pole with an input signal V_k and an output signal V_u , which shows certain storage properties. For our purpose this four pole serves as a system element suitable for use in transmission systems with bandwidth compression in which storage of information is required. The relation between output signal and input signal of this system element now will be derived.

Consider a small area of the target with the dimensions h and dx . The potential of this target element is assumed to be constant over its area $h dx$. The potential difference between target element and the cathode will be denoted by V . The shape of the beamspot is assumed to be rectangular and to have the dimensions h and l . Let V be equal to V_b at the instant that the spot reaches the target element. See fig. 4.

When ρ is the signal current density and C the capacitance of the target per cm^2 then

$$CdV = -\rho dt. \quad (2)$$

Now in a certain range of values for V an exponential relation exists between ρ and V :

$$\rho = \rho_0 \exp(bV) \quad (3)$$

where ρ_0 is a constant depending on the electron-gun grid voltage and b is a function of cathode temperature γ). By integrating and substituting $V = V_b$ at $t = 0$, we find

$$\exp(-bV) = \rho_0 b t / C + \exp(-bV_b). \quad (4)$$

At $t = l/v$ the target element has been completely scanned (v being the scanning velocity). At this moment V is assumed to be V_e and

$$\exp(-bV_e) = \rho_0 b l / C v + \exp(-bV_b)$$

$$\text{or} \quad \exp(-bV_e) = k + \exp(-bV_b), \quad (5)$$

$$\text{where} \quad k \equiv \rho_0 b l / C v. \quad (6)$$

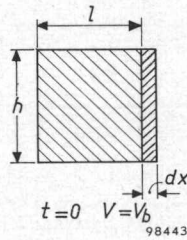


Fig. 4. The rectangular spot $h l$ just before scanning the target element $h dx$. ($t = 0$, $V = V_b$)

Now in a time interval dt an area has been scanned which is $h v dt$. The charge taken up by the area is

$$q = (V_e - V_b) Ch v dt.$$

Hence the signal current can be written as

$$-dq/dt = i = -(V_e - V_b) Ch v.$$

When Chv , which is the capacity scanned per second, is denoted as C_s , so

$$C_s \equiv Ch v, \quad (7)$$

we get

$$i = C_s(V_b - V_e). \quad (8)$$

Substitution of (8) in (5) gives

$$\exp(pi) = k \exp(bV_b) + 1 \quad (9)$$

or

$$\exp(-pi) = -k \exp(bV_e) + 1 \quad (10)$$

in which the new symbol

$$p \equiv b/C_s. \quad (11)$$

In between two successive scans of a target element the target potential V will increase owing to the discharge of the target element capacitance through its resistance. The resistance of the target element is R/dO , in which dO is the area of the element. After one frame period a new value of V_b will be found which in general will differ from the preceding one. Therefore in future calculations V_e and V_b will have to be given an index denoting the serial number of scan; in (9) and (10) V_b , V_e and i must be written $V_{b_{n-1}}$, $V_{e_{n-1}}$ and i_{n-1} . We shall now work out the relation between V_{b_n} and $V_{e_{n-1}}$.

Just after the $(n-1)$ th scan the voltage across the target element is

$$(V_p - i_{n-1}R_s) - (V_{e_{n-1}} + V_{k_{n-1}}).$$

At the beginning of the n th scan, one frame period T later, this has been decayed to

$$\exp(-T/RC) \{ (V_p - i_{n-1}R_s) - (V_{e_{n-1}} + V_{k_{n-1}}) \}$$

At the gun side of the target, therefore, the potential is at the beginning of the n th scan

$$(V_p - i_n R_s) - \exp(-T/RC) \{ (V_p - i_{n-1} R_s) - (V_{e_{n-1}} + V_{k_{n-1}}) \}$$

which is identical to $V_{b_n} + V_{k_n}$.

Therefore

$$V_{b_n} = -\Delta V_{k_n} - \Delta i_n R_s + \alpha V_p - \alpha V_{k_{n-1}} - \alpha i_{n-1} R_s + V_{e_{n-1}}(1 - \alpha) \quad (12)$$

where

$$\exp(-T/RC) = 1 - \alpha \quad (13)$$

and

$$V_{k_n} - V_{k_{n-1}} = \Delta V_{k_n}, \text{ etc.}$$

The foregoing is based on the assumption that the photo current through the target is proportional to the potential difference across it. However, in certain circumstances the current may be less dependent on the voltage and may even be a constant. This is so if the voltage across the target, or the life-time of the charge-carriers is so large, that all charge-carriers contribute to the current without recombining. In this case the increase in voltage from one scan to the next is a constant, written here as δV (so the photocurrent being $C_s \delta V$), and V_{b_n} then becomes

$$V_{b_n} = -\Delta V_{k_n} - \Delta i_n R_s + \delta V + V_{e_{n-1}}. \quad (14)$$

In most cases the voltage drop across R_s , $i_n R_s$, has such a small value that this term may be neglected. This can be shown in the following way.

The magnitude of $i_n R_s$ is $R_s C_s (V_b - V_e)$. Now $C_s \approx C_a/T$, where C_a is the capacitance of the whole target area, or about 1000 to 2000 pF. R_s is about 200 k Ω . Hence

$$R_s C_s = R_s C_a / T \approx 2.10^5.2.10^{-9}.25 = 10^{-2}.$$

Neglecting R_s we get instead of (12)

$$V_{b_n} - V_{e_{n-1}} = -\Delta V_{k_n} + \alpha V_p - \alpha V_{k_{n-1}} - \alpha V_{e_{n-1}} \quad (15)$$

and instead of (14)

$$V_{b_n} - V_{e_{n-1}} = -\Delta V_{k_n} + \delta V. \quad (16)$$

By using in (9) the value of V_{b_n} as given by (15) and by using (10) to eliminate $V_{e_{n-1}}$, we obtain

$$\exp(pi_n) - 1 = k^a \exp(b\alpha V_p - b\alpha V_{k_{n-1}} - b\Delta V_{k_n}) \{1 - \exp(-pi_{n-1})\}^{1-a}. \quad (17)$$

So in this way a recurrent relation for $\exp(pi_n)$ is obtained.

Introducing

$$\beta_n \equiv k^a \exp(b\alpha V_p - b\alpha V_{k_{n-1}} - b\Delta V_{k_n}) \quad (18)$$

we can write

$$\exp(pi_n) - 1 = \beta_n \{1 - \exp(-pi_{n-1})\}^{1-a} \quad (19)$$

When (16) is applied we get

$$\exp(pi_n) - 1 = \beta_n \{1 - \exp(-pi_{n-1})\} \quad (20)$$

with

$$\beta_n \equiv \exp(b\delta V - b\Delta V_{k_n}). \quad (21)$$

From (17) and (20) it appears that in principle the result depends on all earlier values of β and hence on all forgoing values of V_k . Thus the output signal during the n th scan contains information about all preceding frame periods and not only about the n th and $(n-1)$ th frame periods. Bearing in mind our aim, to make available the information in the previous frame period, we have to ask ourselves under what circumstances the contribution of the $(n-2)$ th and earlier frame periods may be neglected.

When (20) applies a rather simple expression for $\exp(pi_n)$ can be given⁸⁾. It can be shown that in this case

$$\exp(pi_n) = 1 + \left[\frac{1}{\beta_n} + \frac{1}{\beta_n \beta_{n-1}} + \dots + \frac{1}{\beta_n \dots \beta_1 \{ \exp(pi_0) - 1 \}} \right]^{-1}. \quad (22)$$

If δV is sufficiently large with respect to ΔV_k , that is if β_n is always sufficiently large, $\exp(pi_n)$ will be equal to β_n :

$$\exp(pi_n) = \beta_n. \quad (23)$$

This approximation is fairly good if β_n is always greater than 20, say. In this case

$$\frac{1}{\beta_n} < \frac{1}{\beta_n} + \frac{1}{\beta_n \beta_{n-1}} + \dots < \frac{1}{\beta_n (1 - 1/20)} = 1.053 \frac{1}{\beta_n}$$

and therefore the error is less than 6%. It follows from (19) that

$$\beta_n > 20 \quad \text{if} \quad b(\delta V + \Delta V_k) > 3.$$

In the vidicons used b was about 5. Hence

$$\beta_n > 20 \quad \text{if} \quad \delta V + \Delta V_k > 0.6 \text{ volt.}$$

Thus if the potential variation due to the photocurrent is only 0.6 V more than the maximum value of ΔV_k , approximation (25) may be employed.

With $\exp(pi_n) = \beta_n$ and $pi_n = b(\delta V + \Delta V_k)$ it follows that

$$i_n = C_s(\delta V + \Delta V_{k_n}). \quad (24)$$

If the d.c. component $C_s\delta V$ — equal to the photocurrent — is disregarded, and if V_p is constant, we get

$$i_n = C_s(V_{k_n} - V_{k_{n-1}}) \quad (25)$$

which is in fact proportional to the difference between the signals of two succeeding frames.

It should be noted that under the above conditions δV , and hence the photocurrent value, does not affect the output signal level.

Of course the relation (24) itself is not a very surprising result. It could have been derived as well by a brief physical description of the scanning process which has been mentioned in section III.2. However, the calculation has been carried out in order to determine quantitatively the conditions for the validity of the expression (24) and the effects occurring when these conditions are not fulfilled. Moreover the result involves data which are applicable when a more complex memory function is contemplated. (This may be a storage action lasting more than two frames, being necessary when one writing scan must be read several times.) In fact all these data can be obtained from (22).

As mentioned before, in this study we go into more detail only for the application in duplex transmission. When outlining the practical arrangements, for duplex transmission, we supposed that video information would only be present during alternate frame periods. Accordingly, V_k alternates between the video-signal value for the target element concerned and a d.c. voltage. When the image is stationary only two values of V_k occur for a given target element, V_{k1} (d.c. voltage) and V_{k2} (video signal); β and i likewise alternate between two values, β_1 or β_2 and i_1 or i_2 .

From (20) it follows that

$$\exp(pi_1) = \frac{\exp(-b\Delta V_k) + \exp(b\delta V)}{\exp(-b\Delta V_k) + \exp(-b\delta V)} \quad (26)$$

and

$$\exp(pi_2) = \frac{\exp(b\Delta V_k) + \exp(b\delta V)}{\exp(b\Delta V_k) + \exp(-b\delta V)} \quad (27)$$

in which

$$\Delta V_k = V_{k2} - V_{k1}. \quad (28)$$

As V_{k1} is the same for all target elements this value can be assumed to be zero. This gives

$$\exp(p_{i1}) = \frac{\exp(-bV_{k2}) + \exp(b\delta V)}{\exp(-bV_{k2}) + \exp(-b\delta V)} \quad (29)$$

and

$$\exp(p_{i2}) = \frac{\exp(bV_{k2}) + \exp(b\delta V)}{\exp(bV_{k2}) + \exp(-b\delta V)}. \quad (30)$$

In fig. 5, $y = p_{i1}/b\delta V = i_1/C_s\delta V$ is shown as a function of $x = V_{k2}/\delta V$ for various values of $b\delta V$. It can be seen that for $b\delta V = 5$ the characteristic is linear over almost all its range between the asymptotes. For larger values of $b\delta V$ the transitions to the asymptotes are still more sharp. The centre of the range in which a linear relationship exists between i and V_{k2} is at $V_{k2} = 0$. It follows that, for the most efficient use to be made of this range, the cathode signal will have to be symmetrical about the zero line, i.e., the variations in V_k during even frame periods will have to be symmetrical about its constant level during the odd ones.

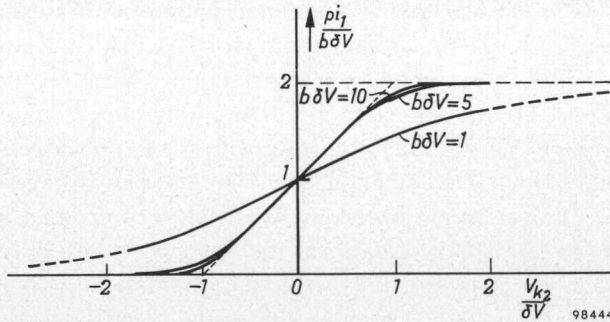


Fig. 5. The output $p_{i1}/b\delta V = i_1/C_s\delta V$ during reading as a function of the input $V_{k2}/\delta V$ during writing for the arrangement of fig. 4 when in alternating frames the input voltage is V_{k2} (writing) and zero (reading). The photo current is assumed to be voltage independent and is equal to $C_s\delta V$.

Of course the image is not usually stationary. In this case, too large a peak value of the cathode signal will affect picture quality owing to non-linear distortion as well as owing to the storage action lasting more than two frames. The occurrence of this storage action can be seen from eq. (22). This will cause a certain smearing effect if the picture content changes suddenly. Conversely, as long as non-linear distortion is avoided by limiting the peak value of the cathode signal, also the memory function will be correct.

When the relation between i_n and i_{n-1} is given by (19) instead of (20) it is very difficult to find a general solution for i_n . This is also true of particular values of α , for instance $\alpha = \frac{1}{2}$. However, it will be clear that i_n will be prac-

tically dependent only on β_n if $\exp(pi_{n-1})$ is sufficiently large. If also β_n is sufficiently large pi_n will approximate $\ln \beta_n$. See eq. (19). Since

$$\beta_n = k^a \exp(baV_p - baV_{k_{n-1}} - b\Delta V_{k_n})$$

it follows that

$$i_n = C_s \{a \ln k/b + aV_p + (1-a)V_{k_{n-1}} - V_{k_n}\}$$

or

$$i_n = C_s \{(1-a)V_{k_{n-1}} - V_{k_n}\} \quad (31)$$

if the d.c. component is neglected.

In the output signal, therefore, the signal of the preceding frame is weaker by a factor of $(1-a)$ than that of the frame now being scanned.

Approximation (31) will be applicable if $a \ln k + baV_p$ is sufficiently large compared to the peak value of V_k , and this depends on the value of $k = \rho_0 bI/Cv$. So in this case the proper functioning depends on the photo current as well as on the beam current. In the forgoing consideration this has not been found.

Now in practice the beam-current value is always important in that it must be above a certain limit. This is due to the fact that the relation $\rho = \rho_0 \exp(bV)$ only applies over a limited range of V . Beyond this range a saturation of ρ occurs. For a smaller value of ρ_0 this range is also decreased and the saturation point is reached more quickly. As shown above, however, ρ_0 has some influence even in the exponential portion of the curve when the photo current depends on voltage.

For a stationary picture, where V_k is equal to V_{k2} and zero during alternate frame periods, from (19) we can derive

$$\{\exp(pi_1) - 1\}^{-1} = \frac{1}{\beta_1} \left\{ 1 + \frac{1}{\beta_2} \left(\frac{\exp(pi_1)}{\exp(pi_1) - 1} \right)^{1-a} \right\}^{1-a}. \quad (32)$$

A similar formula holds for i_2 .

Now i_1 can be found as a function of V_{k2} . For the special case $a = \frac{1}{2}$ the relation is expressed in fig. 6 with

$$a \ln k + baV_p = 5.$$

For comparison, the curve of fig. 5 has been drawn for the same value of $b\delta V$.

Curve I is not symmetrical about points $V_{k2} = 0$. The reason will be clear if it is borne in mind that in this case the photo current has been assumed to be dependent on the voltage across the target. For negative values of V_{k2} the mean value of this voltage across the target will be larger than for positive values of V_{k2} . So also will the photo current, and consequently the voltage increase across the target element in the interval between two scans. This increase of voltage determines the peak value of the cathode signal under conditions of linear behaviour.

From the calculation given above it follows that conditions can be created such that a linear relationship always exists between in- and output signal and that the output signal is only related to the difference of two succeeding frame signals.

It is possible to get a composite output signal V_u by superimposing, with proper amplitude and polarity, the input signal on V_p . In this way, $V_{k_n} + V_{k_{n-1}}$ or $V_{k_{n-1}}$ can be obtained as output signal instead of $V_{k_n} - V_{k_{n-1}}$.

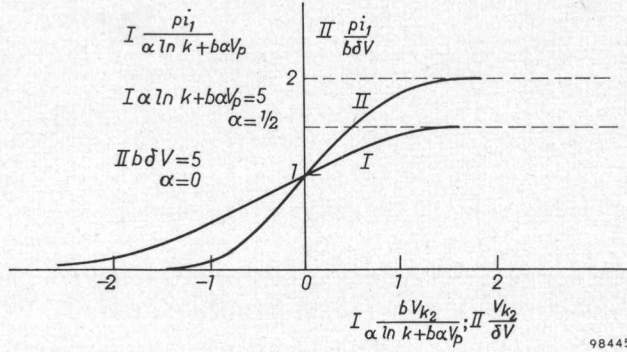


Fig. 6. The output $p_{i1}/(\alpha \ln k + b\alpha V_p)$ during reading as a function of the input $bV_{k2}/(\alpha \ln k + b\alpha V_p)$ for the arrangement of fig. 4 when in alternate frames the input voltage is V_{k2} and zero. I. The photo current is proportional to the voltage across the target; $\alpha = 1 - \exp(-T/RC) = \frac{1}{2}$, $\alpha \ln k + b\alpha V_p = 5$. II. The photo current is independent of voltage; $b\delta V = 5$. Input: $p_{i1}/b\delta V = i_1/C_s\delta V$. Output: $V_{k2}/\delta V$.

The main shortcomings to which the output signal may be subject will be too narrow a bandwidth and a modulation of line or field frequency, or both. The first fault is caused by the spot size, which comes into play during both writing and reading. The second fault is caused by non-linear deflection and non-uniformity of the photosensitive layer. Both effects give rise to a variation in C_s , to which the output signal is directly proportional. In the following section this is considered more in detail.

III.2.2. Experiments on duplex transmission

In order to carry out practical investigations about a system of duplex transmission based on the principles given in the preceding section, the arrangement shown in fig. 7 was built up. Frame and line synchronization pulses are supplied by the generator G . In stage T a test signal is generated which may have various waveforms corresponding to screen images of vertical bars, horizontal bars or a gray scale. This test signal is blocked periodically during one or two fields by an electronic switch S_1 . The signal formed in this way is fed to the cathode of a vidicon-type camera tube Vi with a lead-oxide target 7). The usual voltage

and current supplies for focussing, deflection and so on are also applied to the vidicon. The illumination of the target is uniform and can be varied. The output signal of the vidicon is amplified (A) and then fed to a second electronic switch S_2 where several ways for switching and matrixing are possible in order to get the final output signal. This signal is reproduced on the monitor M .

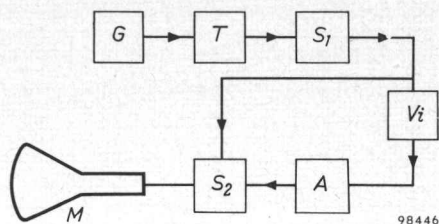


Fig. 7. Arrangement for practical investigation of a frame-sequential duplex transmission system. G = synchronizing generator, T = test-signal generator, S_1 = electronic switch operated at field or frame frequency, S_2 = idem, A = amplifier, M = monitor.

In order to avoid difficulties due to spot size and the line interlace, the first experiments were carried out without line interlace, and hence with half the normal number of lines. Consequently odd and even fields were identical and the input signal was switched field-sequentially.

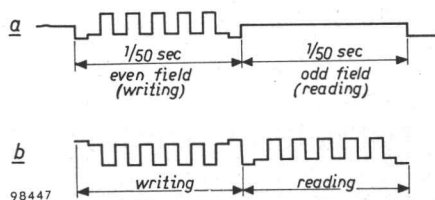


Fig. 8. *a*) Input signal of V_i in fig. 7. In even fields the test signal represents horizontal bars, in odd fields the input is a constant voltage. (Subdivision in line intervals is not drawn.) *b*) Output signal of V_i (idealized) when the signal of 8*a* is applied to the input.

Figure 8*a* shows the input signal for a special test signal (horizontal bars). The d.c. level during odd fields can be chosen more or less arbitrarily but in view of the results obtained in the preceding section (fig. 6), the d.c. level was fixed about midway between the peak values of the video signals. The output signal of the vidicon appears in fig. 8*b*, in ideal form. Signals in writing and reading intervals are of opposite polarity; in both intervals the signal range is the same.

In practice several deviations will occur. These deviations are caused by:

- (1) the spot size, which decreases the resolution during writing as well as during reading;

- (2) the non-linearity of the deflection, which results in a non-uniform scanning speed, introducing variations of the output signal in line and field frequencies because the output is proportional to the scanning speed, as follows from (7), (25) and (31);
- (3) the non-uniformity of the target surface also leading to variations in signal level;
- (4) the fact that the electron beam does not strike the target perpendicularly over the whole area, resulting in a deviation in formula (3);
- (5) leak currents between adjacent target elements which affect the charge pattern, especially wiping out picture detail.

Finally, the capacitance between adjacent target elements may have some influence.

We shall not consider all these causes in detail. We shall only describe measures for decreasing the two main effects, loss of resolution and the occurrence of a signal distortion of line and field frequencies.

Loss of resolution is mainly due to the spot size. In the horizontal direction it can be largely compensated by frequency correction of the input and the output signals. Of course, an improvement may also be made by diminishing the relative spot size, that is to say by widening the area scanned.

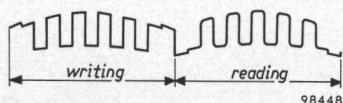


Fig. 9. Output signal of fig. 8b with distortion due to non-uniformity of the scanning process over the target area and due to the finite spot size.

The signal distortion of line and field frequencies appears as a spurious component simply added to the output signal and also in the form of modulation of the output signal. See fig. 9. The former is due to the presence of a d.c. component in the signal current. To some extent the component with field frequency can be diminished by clamping the signal to a fixed level at the beginning of each line period. In addition we cut down the effect by making special arrangements for modifying the shape of the deflection currents. In one or more non-linear, variable stages the deflection waveform were modified in such a way and to such an extent that the resulting variations of scanning speed gave optimum compensation. It was possible by this means to reduce the effect considerably.

The tolerances for this distortion described above depend on threshold values in flicker perception, because the effect will cause a certain inequality of the signal in writing and reading intervals and hence of the brightness of successive

frames in the screen image. From the investigations carried out into flicker phenomena by De Lange ⁹⁾ enough data are available to allow these requirements to be worked out, provided brightness and phosphor decay time are known.

For a maximum value of brightness (30 footlamberts) and normal values of phosphor decay times ($\approx 10^{-4}$ sec) the tolerance can be calculated to be about 2% for 25 cycles per second, that is to say for 25 writing and 25 reading intervals per second, as was the case in our first experiment.

The similarity which can be achieved in even and uneven frame periods also depends on the method used in stage S_2 of fig. 7 to produce the monitor signal. This can be done in several ways. During reading, the output video signal has a polarity opposite to that which it has during writing. One method of reversing this polarity is to add the input signal to the output signal with an amplitude twice that of the output signal. See fig. 10 where a is the output signal, b is the input signal and c represents $a + b$. The advantage of this method is that the distortion of the d.c. component is now identical in writing and reading periods. However, the modulation of the signal has different polarities in the two intervals.

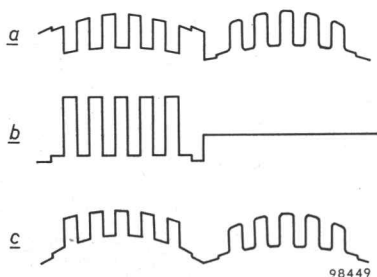


Fig. 10. The display signal c formed by the addition of the output signal a and the input signal b . The d.c. component is distorted at both the line- and field frequency, (only the latter is illustrated here). The distortion is the same in the writing and reading interval. The modulation of the signal at the line and field frequencies has different polarity in the two intervals.

A second method of reversing the signal polarity during the reading interval is to reverse the polarity by switching. However, if this is done, the blanking intervals will be beyond white level (instead of being blacker-than-black) during either the writing or reading interval. This is because the blanking intervals are introduced during writing as well as during reading by suppressing the electron beam. In the original output signal therefore the blanking intervals are fixed on the same level in both intervals.

A third method is to suppress the output signal during writing and to add the undistorted input signal. This method proved to give the best results, making it possible to reduce deviations to less than 5% over the whole picture area.

Although this value is greater than the one mentioned above the picture could be said to be acceptable provided some ambient illumination was present. A certain difference in sharpness was still present between two successive fields, but the effect was not unduly disturbing.

In a second experiment we investigated to what extent the desired memory action could be obtained when normal scanning with 625 lines and line interlace was employed. The difficulties were very much greater than those involved by the previous experiment. In the first place, requirements as to resolution were twice as stringent. In the second place the writing and reading intervals were twice as long as before. In the third place difficulties arose due to the fact that the scanning pattern was now split up into two fields. The distortion in both fields due to non-uniformity had to be as similar as possible, but on the other hand the scanning and leak currents in both fields had to be separated as much as possible. These latter requirements are more or less contradictory, and the question arises as to which is better, to interlace the two fields or to store them in separate areas, one below the other. When the two field patterns are interlaced it is difficult to separate the scanning process and leak currents for both. When the two field patterns are stored one below the other, the difference in distortion will be great.

Experiments have shown the latter difficulty to be less important; accordingly, the best results were obtained by separating the two field patterns completely. However, sharpness and uniformity were not satisfactory. From data on flicker perception it can be calculated that the signal during writing and reading has to be identical within $1/2\%$, as in this case flicker will occur with a frequency of $12\frac{1}{2}$ alternations per second ⁹⁾. It may well be asked whether it will ever be possible to keep tolerances within such narrow limits. Apart from further improvements in uniformity over the target surface, there are two possible ways of getting better results.

In the first place, feedback may in principle be applied in order to get a closer correspondence between the input and output signals. A practical difficulty is the large discrepancy between the input level at the cathode and the output level at the target resistor. Very special precautions have to be taken in order to avoid instability. In the second place the ill-effects of non-uniformity may be eliminated by turning the input signal to a frequency-modulated signal. In this case the input signal is fed to the vidicon as a frequency-modulated carrier. The process is then much less sensitive to amplitude distortion. This principle is employed in techniques for video-signal recording on tape ¹⁰⁾.

It can be concluded from the experiments just described that in principle a duplex transmission is possible with the circuitry and equipment referred to above. However, in addition a special development is necessary in order to remedy the consequences of non-uniformity and to reduce the relative size of the spot.

The system might usefully be employed in long-distance transmission of the kind required in European exchange programmes. The links that have been established for this purpose could be used for two programmes simultaneously provided that line and frame frequencies of both signals were synchronized.

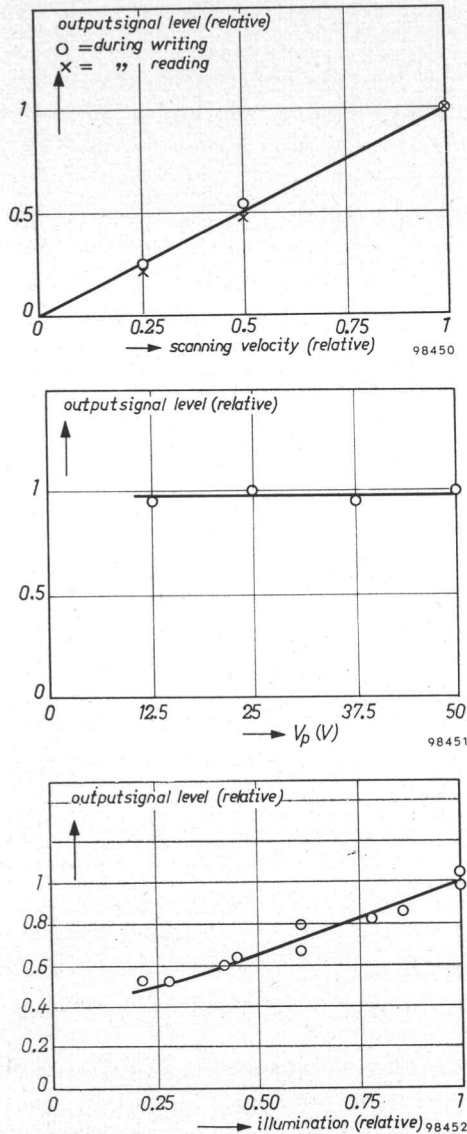


Fig. 11a. Influence of scanning speed on the level of input and output signals of the vidicon V_i in fig. 7.
 11b. Signal level of V_i in fig. 7 at different values of the plate voltage V_p .
 11c. Signal level of V_i in fig. 7 at different illumination levels.

At suitably chosen terminal points the two signals could be separated and passed on via two separate channels.

We shall end this section giving the results of certain measurements that were carried out in the field-sequential experiment. These are shown in fig. 11a, 11b and 11c. Fig. 11a shows the influence of the scanning speed on the signal level. As might have been expected, this level is proportional to the scanning speed, because the charge fed to the target per second is proportional to the area scanned per second. It will be noticed that the signal levels during writing and reading are practically the same. This means that the charge pattern amplitudes written during a certain scan do not decrease within a field interval and are still present in the next reading scan.

Fig. 11b and 11c show the influence of the target potential V_p and the illumination level respectively. Within wide limits the output signal is evidently independent of V_p . From fig. 11c it appears that the output signal is not independent of the illumination level. The influence of V_p and hence of the photo current being of minor importance, it seems reasonable to suspect that the effective target capacitance may vary with the illumination.

III.3. More than two fields per frame

The second possible way of reducing the number of complete frames per second is to decrease the amount of information per vertical scan, the complete picture being broken up into more than two fields. Decreasing the information per field has two meanings here. The reduction can be understood as a mere cutting down of the number of picture elements comprised in one field scanning pattern. Alternatively, the reduction can be affected by adding in a coded form a second signal to the original video signal in such a way that it is of opposite sign in two successive scans of the same area. Due to the persistence of vision the second signal will be barely visible because as far as the eye is concerned its visual effect in one scan cancels out when combined with the visual effect in the successive scan.

We are referring here to the employment of dot-interlace and of subcarriers respectively ¹¹⁾. Both principles were mentioned already in II.3.2. As was stated there, the use of subcarriers is suitable for multiplex transmission, whereas dot-interlace may be employed in the transmission of a single signal with half the normal bandwidth as well as in multiplex transmission requiring a bandwidth of half the added signal bandwidths. As we shall go on to show, there is no sharp difference between the two methods. Yet it is useful to distinguish between dot-interlace systems and subcarrier systems in order to label different aspects of what may be regarded as one principle.

III.3.1. Dot-interlace

We shall first describe the use of dot-interlace and consider the requirements

in such a system for the transmitting process, the transmission-channel characteristics and the receiving process ^{11,12}). As already stated, dot-interlace can be used for single signal transmission or multiplex transmission. In both cases the reduction of channel bandwidth is based on the fact that per scan only half the number of picture elements is transmitted. The distances between the picture elements in one scan is twice as long as normally and therefore only half the normal channel bandwidth is required. In two successive scans of a certain line all picture elements of this line are transmitted, the picture element series of both scans being interlaced. The addition of the two separated series of picture elements will occur at the receiver as a result of the integration effect of the eye.

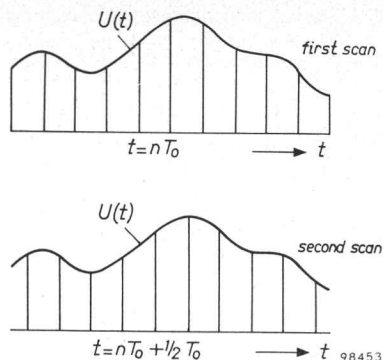


Fig. 12. Principle of dot-interlace. In the first scan the signal values at $t = nT_0$ are separated from the video signal $u(t)$ and transmitted; in the second scan the signal values at $t = nT_0 + \frac{1}{2}T_0$ are transmitted.

For a detailed description of the system let us assume that the interval between picture elements for which values of brightness function $B(x)$ have to be transmitted is $\frac{1}{2}x_0$. In dot-interlace only half the number of picture elements is transmitted per scan and hence each scan comprises a series of values with intervals of x_0 between them. The corresponding time intervals in the corresponding video signal $u(t)$ are assumed to be T_0 . Hence at $t = 0, T_0, 2T_0$ etc. signal values have to be separated from $u(t)$ and fed to the transmission channel. See fig. 12. The transmission path possesses a certain frequency characteristic, and therefore the pulse series will arrive at the receiver with a certain linear distortion. The transmission channel is defined here as the real transmission path together with those stages in the transmitter and receiver which are inserted to affect the frequency band in view of channel bandwidth standards or frequency correction principles.

The receiver has to recover the original signal values, this being done by a sampling process. For this purpose the right signal values are required to occur in the received signal at $t = 0, T_0, 2T_0 \dots$, or rather at instants following these

times with a certain constant time delay. This condition is satisfied if the response of the transmission channel to a special pulse shows zero crossings at the instants at which the succeeding pulses will occur. If this is so, at $t = nT_0$ only the n th signal value corresponding to this moment will contribute to the received signal.

Multiplex transmission is subject to a similar condition regarding pulse response. Then pulses of other signals are inserted between two successive pulses of the same signal, and zero crossings have to occur at corresponding shorter time intervals.

If the sampling process at the transmitter is carried out by infinitely short pulses, the above mentioned condition would be fulfilled by a transmission channel having a rectangular amplitude characteristic and a linear phase characteristic. See fig. 13. As is well known, the unit-pulse response of such a filter is proportional to

$$\frac{\sin 2\pi f_c t}{t} \quad (33)$$

where f_c is the bandwidth.

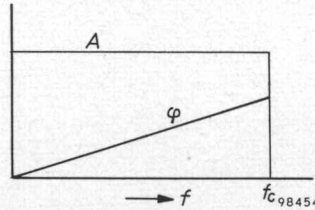


Fig. 13. Frequency characteristic of the ideal filter with response (33) satisfying conditions for independent transmission of successive samples.

Expression (33) is zero at $t = n/2f_c$ ($n = \text{an integer}$), except for $n = 0$. If

$$f_c = 1/2 T_0 = f_0/2 \quad (34)$$

zero crossings occur at the required intervals. See fig. 14.

For the multiplex transmission of k signals, f_c has to be

$$f_c = k/2T_0 = kf_0/2 \quad (35)$$

in which case the zero interval is T_0/k .

At the receiver the appropriate signal values have to be recovered from the received signal by a sampling process. They have to be reproduced at the proper places in the picture area. For the first scan these places are denoted by

$$x = nx_0$$

corresponding to $t = nT_0$.

For the next scan of the same lines, i.e. the third field, the same conditions with respect to the independent transmission of signal samples apply to $t = \frac{1}{2}T_0 + nT_0$ corresponding to $x = \frac{1}{2}x_0 + nx_0$.

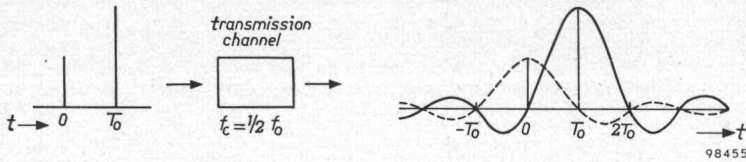


Fig. 14. The transmission of successive samples independent of each other through the filter of fig. 13.

However, as it must be possible to reproduce the brightness of these picture elements quite independently of what has happened in the previous scan, zero brightness must have been reproduced at $x = \frac{1}{2}x_0 + nx_0$ during this previous scan. This condition resembles that to which the unit-pulse response of the transmission channel is subject. If the unit-pulse response of the network inserted between the receiver sampling and the display shows zeroes at the proper time intervals the two interlaced series of picture elements will be reproduced independently. For this network the zeroes have to occur of time intervals of $\frac{1}{2}T_0$. Hence $f_c = f_0$. If sampling were carried out with infinitely short pulses and a filter used as shown in fig. 13 ($f_c = f_0$) the conditions would be satisfied. See fig. 15.

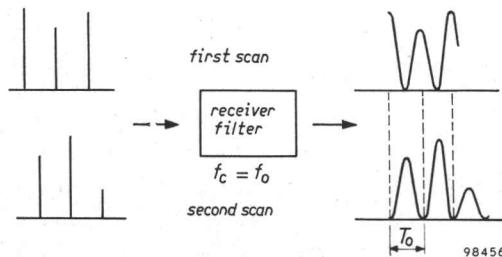


Fig. 15. Receiver sampling and filtering. The samples in two successive scans can be reproduced independently. The addition of the displays of two successive scans reproduces the original video signal, (if the picture is a stationary one).

It will be clear that the employment of infinitely short pulses and ideal filters as in fig. 13 is far from being a practical proposition. We shall therefore go on to investigate whether the requirements can be stated in more general terms, and see to what extent they can be satisfied in practice.

We can reason as follows with regard to the sampling process. The sampling function $h(t)$ for infinitely short pulses can be written

$$h(t) = \lim_{m \rightarrow \infty} \frac{1}{2m+1} \left(1 + \sum_{n=1}^m 2 \cos n\omega_0 t \right) \quad (36)$$

where $\omega_0 = 2\pi f_0$.

Hence the input signal of the transmission channel is

$$u(t) h(t).$$

Under ideal circumstances the frequency range of the transmission path would extend to $\frac{1}{2}f_0$ according to (34). In practice the upper limit will be beyond this value, say f_c . This frequency has not to be understood as the so-called 6 dB point but rather as an upper limit to the passband, beyond which no significant contribution to the transmission process is made. Since the endeavour will be to keep the passband as narrow as possible, frequency f_c may be assumed to be below f_0 , (so twice the theoretical value.)

Now, if the highest frequency in $u(t)$ is f_b , only those terms in (36) will contribute to the output signal of the transmission channel for which

$$nf_0 - f_b < f_0. \quad (37)$$

Since it has been decided that the time interval between picture elements will be $\frac{1}{2} T_0$, apparently only the video signal frequencies up to f_0 have to be transmitted. Therefore it can be assumed that $f_b \leq f_0$. Referring to (37), we see that only the first two terms of (36) are required for the transmission. Accordingly, $h(t)$ can be simplified to

$$h(t) = \frac{1}{3} (1 + 2 \cos \omega_0 t). \quad (38)$$

Similarly, it can be proved that for the receiver sampling the sampling function

$$h(t) = \frac{1}{5} (1 + 2 \cos \omega_0 t + 2 \cos 2\omega_0 t) \quad (39)$$

is satisfactory. This applies only if the transmission characteristic (6dB point at $\frac{1}{2}f_0$) and the receiver filter (6 dB point at f_0) have a sufficiently sharp cut off.

For multiplex transmission of k signals, each with an upper frequency of f_0 , a corresponding simplification of the sampling functions can be made. Here too, the number of terms which have to be taken account of depends on the steepness of the transmission- and receiver-filter characteristics. Under idealised circumstances the channelband extends to $kf_0/2$, the receiver filter to f_0 . That being so, the following sampling function is required

$$h(t) = \frac{1}{2m+1} \left\{ 1 + \sum_{n=1}^m 2 \cos n\omega_0 t \right\} \quad (40)$$

where $m = k/2 + 1/2$ if k is odd and $m = k/2$ if k is even.

From the foregoing it can be seen that the sampling process can be simplified without any prejudicing of the ultimate result to A.M. modulation of a limited number of carriers $f_0, 2f_0 \dots$

With regard to the frequency characteristics it can be shown in the following way that ideal filtering can be avoided. The Fourier transform of a unit-pulse can be written

$$\text{Re} \left[\int_0^{\infty} e^{j\omega t} d\omega \right].$$

The response of an ideal filter with bandwidth $\frac{1}{2}f_0$ and zero phaseshift can be written

$$\text{Re} \left[\int_0^{\frac{1}{2}\omega_0} e^{j\omega t} d\omega \right].$$

A network with frequency characteristic $\overline{W}(\omega)$ and a passband up to f_0 will have the response

$$\text{Re} \left[\int_0^{\omega_0} \overline{W}(\omega) e^{j\omega t} d\omega \right].$$

The integral can be written as

$$\int_0^{\frac{1}{2}\omega_0} e^{j\omega t} d\omega + \int_0^{\frac{1}{2}\omega_0} (\overline{W} - 1) e^{j\omega t} d\omega + \int_{\frac{1}{2}\omega_0}^{\omega_0} \overline{W} e^{j\omega t} d\omega.$$

On the substitution of $\omega = \omega_0 - \omega'$ the latter integral becomes

$$\int_0^{\frac{1}{2}\omega_0} \overline{W}(\omega_0 - \omega) e^{j\omega_0 t - j\omega t} d\omega.$$

Now, at $t = nT_0$ the response has to be zero. The first term is zero at these instants because it is the response of an ideal filter. In the third term,

$$\int_0^{\frac{1}{2}\omega_0} \overline{W}(\omega_0 - \omega) e^{j\omega_0 t - j\omega t} d\omega,$$

$e^{j\omega_0 t}$ can be omitted because

$$n\omega_0 T_0 = n2\pi.$$

Further, since the real part of $\overline{W}(\omega_0 - \omega) e^{-j\omega t}$ equals the real part of $\overline{W}^*(\omega_0 - \omega) e^{j\omega t}$ the condition can be written as

$$\text{Re} \left[\int_0^{\frac{1}{2}\omega_0} \overline{W}(\omega) + \overline{W}^*(\omega_0 - \omega) - 1 \right] e^{j\omega t} d\omega = 0$$

for all values $t = nT_0$. This condition is satisfied if

$$\overline{W}(\omega) + \overline{W}^*(\omega_0 - \omega) = \text{constant}$$

for all values $0 < \omega < \frac{1}{2}\omega_0$. As a constant delay throughout the whole frequency range can be allowed, the condition can be expressed in the following more general form:

$$\overline{W}(\omega) + \overline{W}^*(\omega_0 - \omega) = C \exp(-j\omega\tau) \quad (41)$$

in which C and τ are constants having arbitrary values. Figure 16 represents a special case. Here the phase characteristic is linear and the amplitude characteristic is antisymmetrical with respect to $\frac{1}{2}f_0$.

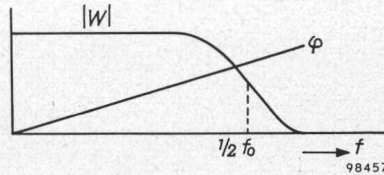


Fig. 16. A special case of a filter characteristic satisfying the condition (41) for an error-free transmission.

From the foregoing it should be clear that the conditions for frequency characteristics can be stated in a general form allowing practical realization. The reasoning was based on the assumption that frequencies beyond f_0 could be neglected. When $\overline{W}(\omega)$ extends so far that frequencies beyond f_0 have to be taken into consideration, condition (41) can be still further generalized. However, we are concerned here with frequency limitation, and the expansion of (41) is therefore of little interest.

III.3.1.1. Realization of the transmission characteristic

Having stated the requirements for dot-interlace transmission in the previous section in a mathematical form, we shall now see to what extent the transmission characteristic can be made to satisfy them in practice. In the first place we shall show how an approximation to (41) can be obtained. As is well known, by the use of lumped inductances and capacitances a network can be formed giving constant attenuation and group delay over a certain frequency range. By employing a sufficiently large number of circuit elements and a sufficiently low impedance level, it is possible to obtain a sufficiently close approximation to such a delay line for a given frequency range and a given delay time. For that frequency band the characteristic can be given as

$$\overline{W} = C \exp(-j\omega t_1)$$

in which t_1 is the time delay. If two such networks, giving delays of $t_1 - t_2$ and $t_1 + t_2$, be connected in parallel, we shall have

$$C \exp\{-j\omega(t_1 + t_2)\} + C \exp\{-j\omega(t_1 - t_2)\} = 2C \cos \omega t_2 \exp(-j\omega t_1). \quad (42)$$

A transfer function proportional to

$$C_n \cos n\omega t_2 \exp(-j\omega t_1) \quad (43)$$

can likewise be arrived at by connecting in parallel two delays $t_1 + nt_2$ and $t_1 - nt_2$.

It will be clear that a series of terms can be combined to give a frequency characteristic with a linear phase shift and a periodic amplitude characteristic, the period being $1/t_2$. More particularly, it is possible to write down, in the form of a progression, an amplitude characteristic of rectangular shape and of period $2f_0$. The first three terms are

$$\left\{ \frac{1}{2} + \frac{2}{\pi} \cos \frac{\pi\omega}{\omega_0} - \frac{2}{3\pi} \cos 3 \frac{\pi\omega}{\omega_0} \right\} \exp(-j\omega t_1). \quad (44)$$

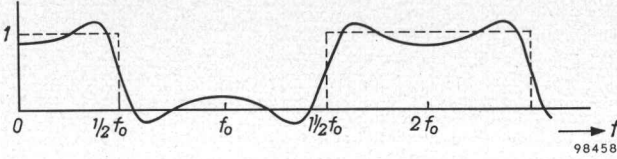


Fig. 17. Periodic filter characteristic satisfying (41) constructed by the use of delay lines.

$$W = \left\{ \frac{1}{2} + \frac{2}{\pi} \cos \frac{\pi\omega}{\omega_0} - \frac{2}{3\pi} \cos 3 \frac{\pi\omega}{\omega_0} \right\} \exp(-j\omega t_1).$$

The corresponding curve appears in fig. 17. This form is such as to satisfy the condition (41):

$$\overline{W}(\omega) + \overline{W}^*(\omega_0 - \omega) = C \exp(-j\omega\tau)$$

but it has no upper frequency limit. However a low-pass filter can be employed for limiting the pass-band. All that this filter is required to do is to provide a linear phase shift from 0 up to $\frac{1}{2}f_0$ and sufficient attenuation at frequencies of $1\frac{1}{2}f_0$ and above.

The low-pass filter can be still further simplified if the period of the characteristic is increased, maintaining the bandwidth equal to $\frac{1}{2}f_0$. See fig. 18. Of course, many more terms are necessary in this case. It has, however, to be borne in mind that increasing the number of terms is going to create difficulties, because the n th term will require a delay line with a delay of $2nt_2$ sec. This can be seen from

$$C_n \exp \{-j\omega(t_1 - nt_2)\}$$

in which only positive values of $t_1 - nt_2$ are possible. Therefore in

$$C_n \exp \{-j\omega(t_1 + nt_2)\}$$

$t_1 + nt_2$ is at least $2nt_2$.

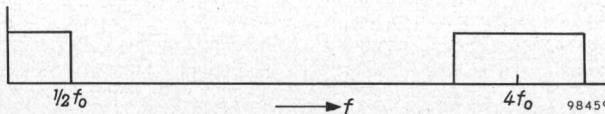


Fig. 18. Periodic filter characteristic with a period $4f_0$ and a low-pass band equal to $\frac{1}{2}f_0$.

The required delay times can be halved if open-ended delay lines are employed, advantage being taken of the reflection at the open end in order to get

$$C_n \cos n\omega t_2 \exp(-j\omega t_1)$$

half-way between the input and output terminals¹³⁾.

The employment of delay lines in the manner just described is not the only possible way of using delays in order to synthesize the required characteristic. This can be illustrated by the following example, in which again the unit-pulse response is considered¹⁴⁾. The filter used is illustrated in fig. 19a. The unit-impulse response of this network is

$$w(t) = \exp(-at) \sin \omega_0 t/2 \quad (45)$$

corresponding to one pair of simple poles at

$$j\omega = -a \pm j\omega_0/2.$$

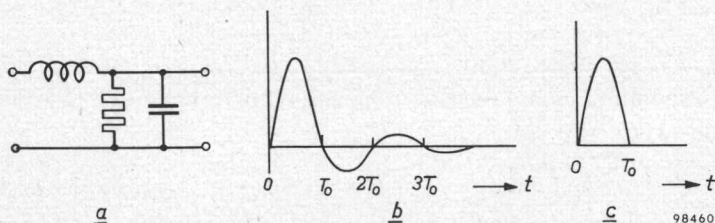


Fig. 19. An example of using delay lines to synthesize the required unit-pulse response.

a) Low-pass filter.

b) Unit-pulse response of a; $\exp(-at) \sin \omega_0 t/2$.

c) The response b with a time delay of T_0 sec and with appropriate attenuation is subtracted from the original b to form c.

The response is given in fig. 19b. Now in fig. 19c this response is combined with the same response with a time delay of $t_1 = T_0$ sec and with appropriate magnitude. It will be clear that the resulting shape may satisfy the conditions because for $t < t_1$ the original response is present and for $t > t_1$

$$w(t) = \exp(-at) \sin \omega_0 t/2 + k \exp\{-a(t - t_1)\} \sin \omega_0(t - t_1)/2 \quad (46)$$

is zero if

$$k = \exp(-at_1). \quad (47)$$

In this construction no attention is paid to the frequency characteristic. Analysis shows that it falls off very slowly, extending far beyond $f_0/2$ and this is not of course in accordance with our aim. However, this example has only been given to illustrate what can in principle be done. By using more filter sections and various time delays a more elaborate synthesis of the pulse response can be obtained, with a sharper cut-off of the frequency characteristic.

Though it has been proved above that the desired transmission characteristic can be approximated as closely as may be desired, the question still remains as to whether the conditions can be fulfilled by a “finite network”, i.e. a network consisting of a finite number of lumped constants. To consider this problem we shall describe the desired response in a general way.

The response of the network to a unit-pulse at the instant $t = 0$ has to show zeroes at $t = t_1 + nT_0$ for all integer values of n except one, that being $n = n_s$. Here t_1 has an arbitrarily value which must however be smaller than T_0 . The response at $t_s = t_1 + n_s T_0$ is sometimes called the “characteristic sampling point”. See fig. 20. ($n_s = 2$.)

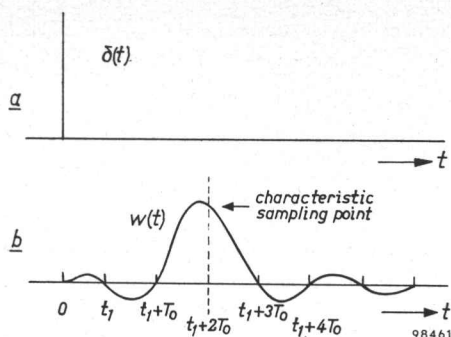


Fig. 20. General presentation of a unit-pulse response satisfying the required conditions
a) Unit-pulse at $t = 0$.
b) Unit-pulse response, zero at $t = t_1 + n T_0$ for all n except $n = 2$. (“characteristic sampling point”).

The general expression which satisfies these conditions is

$$w(t) = \frac{A(t)}{t - t_s} \sin \frac{1}{2} \omega_0 (t - t_s) \quad (48)$$

where $A(t)$ is a finite function which is not zero at $t = t_s$. On the other hand the response of a network which comprises a finite number of lumped constants (and which may embody electronic tubes) can be written down in a finite number of terms like

$$t^n \exp(-at) \sin(\omega_p t + \phi)$$

in which n is an integer > -1 and a is positive.

Now it is impossible to identify a finite series of these terms with (48). Hence it is impossible to achieve the desired pulse response exactly with a “finite network”. On the other hand it is of course possible to get a close approximation by adding a number of suitable terms $t^n \exp(-at) \sin(\omega_p t + \phi)$ with appropriate coefficients.

III.3.1.2. The introduction of a third signal process

The system of transmission so far described comprised two processes, sampling and filtering, each of these having its special conditions. However some examples can be given of a third process to which the signal may be subjected, with a view to simplify network requirements. In the first place it should be noted that the time delays mentioned in the preceding section can be obtained by means other than electrical delay lines. For instance, magnetic or electrostatic recording equipment may be used for this purpose. So this practice can be introduced as a third signal process.

In the second place a different pulse shape can be introduced in order to make the network easier to design. Up to now the input signal of the transmission channel has been represented as a series of infinitely short pulses possibly lacking those harmonic frequency components which cannot contribute to the output signal on account of the limited pass-band. Here we shall consider an extension of the pulses in such a way that each unit-pulse gives rise to a pulse of fixed shape and finite width T_p . ($T_p < T_0$.) See fig. 21. It will be noted that this process is different from sampling with pulses of a special form and width because in the latter case multiplication of the video signal and pulses takes place, and the pulse shape in the output signal depends on the video signal. In fig. 21 the pulse shape is fixed, only its height being determined by the video signal. It is a fairly easy matter to make electronic devices perform the process represented in fig. 21.

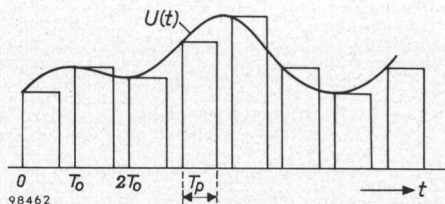


Fig. 21. A third signal process in the time-division transmission. Each sample gives rise to a pulse of fixed shape; only its height is determined by the video signal $u(t)$.

When this process is employed the conditions laid down in fig. 20 for the unit-pulse response $w(t)$ must now be satisfied by the response of the transmission channel to the broadened pulse. If $p(t)$ denotes the shape of the pulse, its response $w_p(t)$ can be written as

$$w_p(t) = \int_0^t p(\tau) w(t - \tau) d\tau \quad \text{when } t < T_p, \quad (49a)$$

$$w_p(t) = \int_0^{T_p} p(\tau) w(t - \tau) d\tau \quad \text{when } t < T_p. \quad (49b)$$

Now zero conditions can be fulfilled by a “finite network” for a suitable choice of the pulse form. This will be evident from the simple example, which follows.

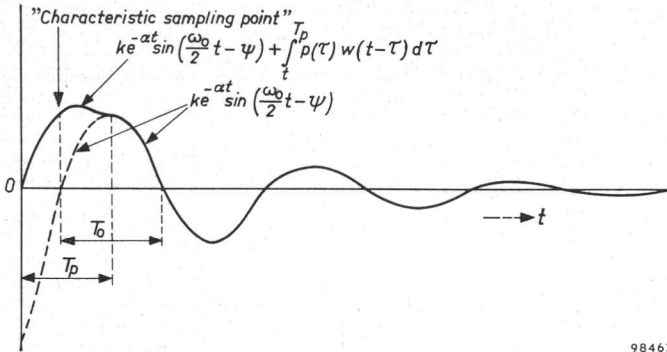
In (49b) the response corresponding to one pair of simple poles is substituted for $w(t)$, hence

$$w(t) = \exp(-\alpha t) \sin \omega_0 t / 2. \quad (50)$$

It can be easily shown that for $t > T_p$, $w_p(t)$, as given by (49b), will have a form similar to $w(t)$, but will involve a constant and a phase shift:

$$w_p(t) = k \exp(-\alpha t) \sin(\omega_0 t / 2 - \psi)$$

which means that outside the interval $0 - T_p$ zero crossings occur at equal intervals of T_0 at instants $t = nT_0 + 2\psi/\omega_0$. See fig. 22.



98463

Fig. 22. The response to a pulse of finite width of a simple low-pass filter corresponding to (49).

Within this interval $w_p(t)$ can be written as

$$\begin{aligned} \int_0^{T_p} p(\tau) w(t-\tau) d\tau - \int_t^{T_p} p(\tau) w(t-\tau) d\tau = \\ = k \exp(-\alpha t) \sin(\omega_0 t / 2 - \psi) - \int_t^{T_p} p(\tau) w(t-\tau) d\tau. \end{aligned} \quad (51)$$

Now, it is possible to get a finite value within the interval at the zero point of $k \exp(-\alpha t) \sin(\omega_0 t / 2 - \psi)$ due to the contribution of the second term in (51), thus fixing a “characteristic sampling point”. The first zeropoint in $k \exp(-\alpha t) \sin(\omega_0 t / 2 - \psi)$ at $t = 2\psi/\omega_0$ will be within the interval when $2\psi/\omega_0 < T_p$. No other point is possible because $T_p < T_0$. Hence a characteristic sampling point will occur if, at $t = 2\psi/\omega_0$

$$\int_t^{T_p} p(\tau) \exp\{-\alpha(t-\tau)\} \sin(\omega_0(t-\tau)/2) d\tau \neq 0. \quad (52)$$

III.3.1.3. Picture reproduction

In this section we propose to describe in more detail how the picture reproduction takes place in a dot-interlace system. Here too we shall consider information in the form $B(x, y, t)$ as has been done in part I. This means that the transformation of the x coordinate in a time scale in the camera and back again in the picture tube is ignored. All statements made about conservation of signal values in a time function will now apply to brightness as a function of x , the horizontal direction. As reproduction in this direction has our special attention here, we shall not consider the dependence of y . Hence the input function of the whole system is written as $B(x, t)$ in which only those values of t at which transmission takes place are of importance. These are denoted by $t = nT_B$ where T_B is twice the field period, i.e. the frame period in normal scanning.

It will be assumed that the sampling process at the transmitter and the transmission characteristic satisfy the conditions that have been laid down. Hence the signal at the receiver input will carry the correct signal values at the instants of sampling. At the receiver these signal values are recovered by a sampling process. As described before in III.3.1 in the sampling function only those terms have to be taken into account which make a contribution to the output of the receiver filter.

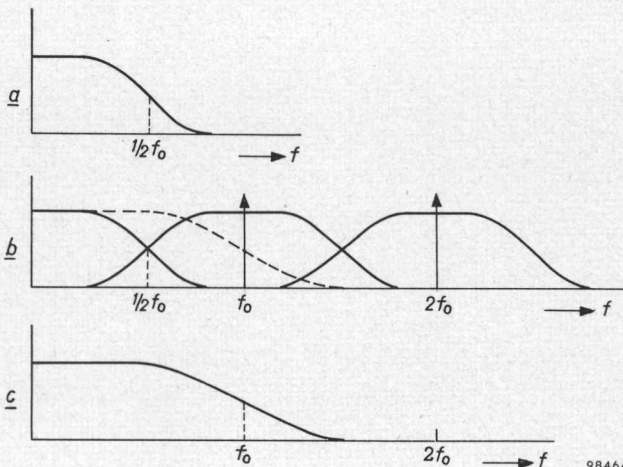


Fig. 23. Frequency-bands occurring when the transmission signal in a dot-interlace system is sampled with

$$h(t) = a_0(1 + 2 \cos \omega_0 t + 2 \cos 2\omega_0 t)$$

at the receiver.

- a) Pass-band of the transmission channel.
- b) The transmission signal multiplied by $h(t)$.
- c) Pass-band of the receiver filter.

As the upper frequency limit of the transmission channel will be below f_0 (nominal bandwidth $\frac{1}{2}f_0$) and the frequency limit of the receiver filter can be assumed to be below $2f_0$ (nominal bandwidth f_0), it can easily be shown that the first three terms of $h(x)$ have to be taken into account. See fig. 23. Hence

$$h(x) = \alpha_0(1 + 2 \cos 2\pi x/x_0 + 2 \cos 4\pi x/x_0).$$

This applies to one particular scan. To the succeeding scan of the same line (i.e. T_B sec later) applies

$$h(x) = \alpha_0(1 - 2 \cos 2\pi x/x_0 + 2 \cos 4\pi x/x_0).$$

In general, for the n th scan

$$h_n(x) = \alpha_0(1 + (-1)^n 2 \cos 2\pi x/x_0 + 2 \cos 4\pi x/x_0). \quad (53)$$

After sampling, the signal is fed to the receiver filter. As was described in III.3.1 the unit-pulse response of this filter has to exhibit zeroes at intervals of $\frac{1}{2}T_0$ and the nominal bandwidth is f_0 . The frequency characteristic has to satisfy the condition

$$\overline{W}(\omega) + \overline{W}^*(2\omega_0 - \omega) = C \exp(-j\omega\tau) \quad (54)$$

as follows from (41). The pulse response $w(t)$ can be transformed into $w(x)$, a function of place, with zero crossings at intervals $\frac{1}{2}x_0$. The output signal of this filter can be represented by

$$\int_{-\infty}^x h(\xi) B(\xi, t) w(x - \xi) d\xi. \quad (55)$$

In this expression $h(\xi)$ is applied to the original function $B(\xi, t)$ instead of a function corresponding to the received signal. This is permissible, because the input signal at the receiver is assumed to carry the original signal values at the instants of sampling, and the ultimate result at the receiver will not differ from that which would have been obtained if this original signal had been fed to the receiver. However, it will only be correct to do so if account is taken of all terms in the sampling function which make a contribution to the output of the receiver filter when the original signal is the input signal. It can easily be shown, that these are the terms also brought into account in (53).

If the influence of the spot diameter, the non-linearity and the decay time of the phosphor in the picture tube are left out of consideration, expression (55) can be regarded as a description of the reproduced image.

In order to determine the visual impression, persistence of vision has to be taken into account. It is assumed that the persistence phenomenon can be described by introducing a unit-pulse response $\exp(-at)$. Now the results of all scans have to be summated up to the m th scan, m being such that $mT_B < t$ and $(m + 1) T_B > t$. This gives

$$\sum_{-\infty}^m e^{-a(t-nT_B)} \int_{-\infty}^x \left(1 + (-1)^n 2 \cos \frac{2\pi\zeta}{x_0} + 2 \cos \frac{4\pi\zeta}{x_0} \right) B(\zeta, nT_B) w(x-\zeta) d\zeta. \quad (56)$$

This expression consists of three terms. The first term represents the impression which exists in normal television transmission, apart from a constant of proportionality, which will be discussed later on. The third term represents a picture component which would also have been present if the sampling frequency had been doubled, so that each scan resulted in the transmission of all picture elements. This term is not very important. It represents the modulation of $B(x, t)$ on $\cos 4\pi x/x_0$. The major part of the corresponding signal will fall outside the frequency range of the receiver filter especially if the attenuation increases sharply beyond f_0 .

The second term is the interfering effect proper, which has to be at as low a level as possible if the dot-interlace is to be successfully used. In virtue of the particular properties that $w(x)$ has, the expression

$$\int_{-\infty}^x \cos \frac{2\pi\zeta}{x_0} B(\zeta, t) w(x-\zeta) d\zeta$$

represents some form of single-sideband modulation of $B(x, t)$ on $\cos 2\pi x/x_0$. This is because $w(x)$ corresponds to a filter with a nominal bandwidth f_0 . Its frequency characteristic satisfies the relation (54). This can be understood

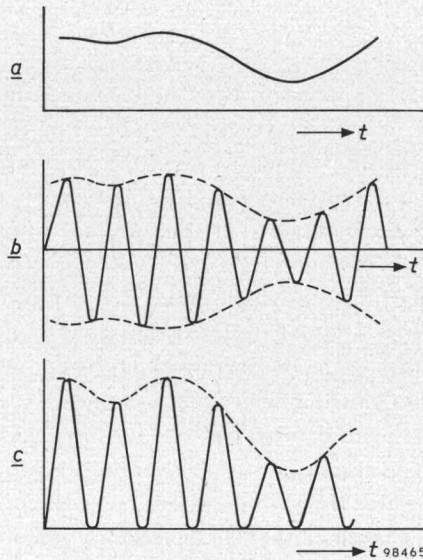


Fig. 24. Presentation of the display signal in an error-free dot-interlace transmission
a) The original video signal. b) The interfering component. c) The display signal, $a + b$, with zero response for half the number of picture elements.

in such a way that the lower sideband ($\overline{W}(\omega)$) and the upper sideband ($\overline{W}(2\omega_0 - \omega)$) — only present very close to the carrier f_0 — added together just deliver the original signal.

It is easy to demonstrate, by reasoning based on the properties of $w(x)$, that the addition of

$$\int_{-\infty}^x B(\zeta, t) w(x - \zeta) d\zeta \quad (\text{fig. 24a})$$

and

$$2 \int_{-\infty}^x \cos \frac{2\pi\zeta}{x_0} B(\zeta, t) w(x - \zeta) d\zeta \quad (\text{fig. 24b})$$

produces a waveform such as shown in fig. 24c. As can be seen the ratio between the peak-to-peak value of the interfering component and that of the desired component is two. This holds for the picture brightness. To know this ratio in the visual impression the summation indicated by (56) has to be carried out.

Of course, the visual impression — i.e. picture quality — can not fully be characterized by the formula (56). The “visual activity” of the observer has to be also taken into account. As already explained in paragraph II.3.2.2, the eye movements of the observer are subjected to certain stroboscopic effects caused by the dot-interlace. The eyes tend to move in such a way that the dot structure of successive fields coincide on the retina and therefore are very well visible. Apart from this effect, the interference in the dot-interlace image can be determined from the above expression. In our opinion a discussion of this interference can be restricted to three special cases, this covering all essential aspects.

In the first case the picture is assumed to be stationary, so that $B(x, t)$ can be replaced by $B(x)$.

In the second case it is assumed that there is only a variation in contrast, so that $B(x, t)$ can be written as

$$B(x, t) = F(t) B(x).$$

In the third case the objects in the picture are assumed to be moving with a speed v in horizontal direction, so that $B(x, t)$ can be written as

$$B(x, t) = B(x - vt).$$

I. $B(x, t) = B(x)$.

Expression (56) becomes

$$\begin{aligned} & \sum_{-\infty}^m e^{-a(t-nT_B)} \int_{-\infty}^x B(\zeta) w(x - \zeta) d\zeta + \\ & + 2 \int_{-\infty}^x \sum_{-\infty}^m (-1)^n e^{-a(t-nT_B)} \cos \frac{2\pi\zeta}{x_0} B(\zeta) w(x - \zeta) d\zeta + \\ & + 2 \int_{-\infty}^x \sum_{-\infty}^m e^{-a(t-nT_B)} \cos \frac{4\pi\zeta}{x_0} B(\zeta) w(x - \zeta) d\zeta. \end{aligned}$$

In the first term the summation

$$\sum_{-\infty}^m e^{-\alpha(t-nT_B)}$$

can be transformed into

$$e^{-\alpha(t-mT_B)} \sum_0^{\infty} e^{-\alpha nT_B} = \frac{\exp\{-\alpha(t-mT_B)\}}{1 - \exp(-\alpha T_B)}. \quad (57)$$

Similarly, the summation in the second term can be written

$$\frac{\exp\{-\alpha(t-mT_B)\} (-1)^m}{1 + \exp(-\alpha T_B)}. \quad (58)$$

Hence due to the persistence of vision the ratio between the unwanted and the desired component, as represented in fig. 24, is decreased by a factor of

$$\frac{1 - \exp(-\alpha T_B)}{1 + \exp(-\alpha T_B)}$$

and becomes

$$(-1)^m \frac{1 - \exp(-\alpha T_B)}{1 + \exp(-\alpha T_B)}. \quad (59)$$

The factor $(-1)^m$ indicates that the interfering effect is of opposite polarity in successive scans, and hence there will be some flicker. However, over large areas the mean interference per scan is zero, and therefore there is only a flickereffect in picture detail.

When persistence of vision is put at 0.1 sec, so α then being 10, and T_B is 1/25 sec the ratio (59) has a value of 0.3.

II. $B(x, t) = F(t) B(x)$.

In this case the time dependence of the picture is only a variation in contrast. Now (56) involves the summation

$$\sum_{-\infty}^m e^{-\alpha(t-nT_B)} F(nT_B) \quad (60)$$

and

$$\sum_{-\infty}^m (-1)^m e^{-(\alpha t - nT_B)} F(nT_B). \quad (61)$$

When the function $F(t)$ is monotonic in a particular interval, the influence of $F(t)$ in (60) and (61) can be interpreted as an increase or decrease in persistence of vision. In the first case the relative amount of interference will decrease, whereas in the second case it will increase. If $F(t)$ falls off sharply enough, the interference may disappear altogether. Beyond this value it will reappear with opposite sign. Of course the variation in $F(t)$ and the corresponding time inter-

val are not independent. The faster the contrast changes the sooner an extreme signal value will be reached. A simple expression can be given for (60) and (61) if $F(t)$ varies linearly or exponentially in time. The influence of $F(t)$ on the interference is small or of very short duration, except in very special circumstances, for instance where the picture content varies periodically from one frame to the other, (Spokes of turning wheels.)

III. $B(x, t) = B(x - vt)$.

The picture content is moving horizontally. The viewer can only observe it closely by following it with his eyes, keeping the image stationary on the retina. It follows from this that in the summation x has to be replaced by $x + vnT_B$. This gives

$$\sum_{-\infty}^m e^{-a(t-nT_B)} \int_{-\infty}^{x+vnT_B} \left\{ 1 + 2(-1)^n \cos \frac{2\pi\zeta}{x_0} + 2 \cos \frac{4\pi\zeta}{x_0} \right\} B(\zeta - vnT_B) w(x + vnT_B - \zeta) d\zeta.$$

By substitution ζ' for $\zeta - vnT_B$ we get the following for the interfering term

$$\sum_{-\infty}^m e^{-a(t-nT_B)} \int_{-\infty}^x 2(-1)^n \cos \frac{2\pi}{x_0} (\zeta' + vnT_B) B(\zeta') w(x - \zeta') d\zeta'. \quad (62)$$

Now

$$\sum_{-\infty}^m e^{-a(t-nT_B)} (-1)^n \cos \frac{2\pi}{x_0} (\zeta' + vnT_B)$$

can be calculated to be proportional to

$$\left\{ 1 + 2 \exp(-aT_B) \cos \frac{2\pi}{x_0} vT_B + \exp(-2aT_B) \right\}^{-\frac{1}{2}}. \quad (63)$$

It can be seen from this expression that this function is periodical in v . If the value of v is close to an odd multiple of $\frac{1}{2}x_0$ per T_B seconds the interference is very strong, being of the same magnitude as the desired signal. The interlace of picture elements is no longer present for the observer because the dot series of two successive scans of a certain line coincide on his retina. On the other hand if the speed of motion equals an even multiple of $\frac{1}{2}x_0$ per T_B seconds the interfering term is as large as in the stationary picture.

For the third term similar effects can be derived which however are not given here because of their comparative unimportance.

Having given this account of the relative interfering effect, we must make a further observation with regard to the absolute level of the desired signal.

The picture tube will only be able to reproduce a certain maximum value of brightness. Generally the peak value of the video signal fed to the picture tube will correspond to this maximum brightness. Now in the display signal of the

dot-interlace system the peak value is twice the peak value of the desired signal component, as can be seen from fig. 24. Hence, in a dot-interlace system with a conventional display, the maximum brightness is half the value it has in a normal television system. Of course, this is a serious disadvantage. The disadvantage could however be eliminated if it were possible to add the visual effect of two successive scans in some way other than by exploiting persistence of vision. In principle, the arrangement described in paragraph III.2.1 can be employed for this purpose. If the addition of two successive scans is made to take place prior to reproduction the stroboscopic effects described in paragraph II.3.2.2 will likewise be eliminated.

III.3.2. Subcarriers

In the treatment of the dot-interlace system, in III.3.1, the analysis was based on the time-division principle, but concluded with a description in terms of modulated sine waves with frequencies of f_0 , $2f_0$ and so on. These sine waves are modulated by video signals and combined with these video signals in a special relationship such that the time-division principle is maintained. We shall now describe multiplex transmission from the other side, starting from the principle of the employment of subcarriers situated within the video-frequency band and modulated by television signals¹¹).

When a sine wave is added to the frequency-band of a television signal u_1 and the phase relation is such that the sine wave is in opposite phases in two successive scans of the same line, the added signal will be scarcely visible because, as far as the eye is concerned, the spurious brightness modulation caused in one scan cancels out when combined with the brightness modulation due to the next scan. When the sine wave is used as a subcarrier and is modulated by a second television signal u_2 , cancellation still occurs if the modulation is identical in successive scans, that is to say, if u_2 shows only slight variations in successive frames. Furthermore, in the visual reproduction of the subcarrier stroboscopic effects will occur that are similar to those present in dot-interlace systems, described in II.3.2, and it is equally possible to decrease the stroboscopic effect by making a suitable choice of phase relations.

The total video signal transmitted can be represented by

$$u_1 + a u_2 \cos \omega_s t \quad (64)$$

where ω_s is the subcarrier frequency and a is a constant. (The expression does not involve the typical phase properties of the subcarrier.)

When this signal is applied to a picture tube it will produce a visible picture corresponding to u_1 because $u_2 \cos \omega_s t$ cancels out in two successive scans. In order to obtain a picture corresponding to u_2 it is necessary to multiply the transmission signal by $\cos \omega_s t$, that is, to use synchronous detection.

The result of doing this can be expressed as

$$u_2 + u_2 \cos 2\omega_s t + (2/a) u_1 \cos \omega_s t.$$

If this signal is filtered in such a way that the signal component $u_2 \cos 2\omega_s t$ is removed, the result is

$$u_2 + (2/a) u_1 \cos \omega_s t. \quad (65)$$

This signal can be used to reproduce u_2 . If the constant a is chosen to be $\sqrt{2}$, the amplitude relation between signals u_1 and u_2 is symmetrical.

Expression (64) is appropriate to a subcarrier with two sidebands. The results will be just the same if u_2 is transmitted with a single sideband: removal of one sideband will not affect the cancellation process, and synchronous detection will preserve the shape of the modulation. Apart from the smaller bandwidth of the modulated subcarrier, single-sideband transmission has also the advantage of giving better signal-to-noise conditions for u_2 at the receiver. The extreme case of single-sideband operation is shown in fig. 25, where the two signals have the same bandwidth f_s , which is equal to the channel bandwidth.

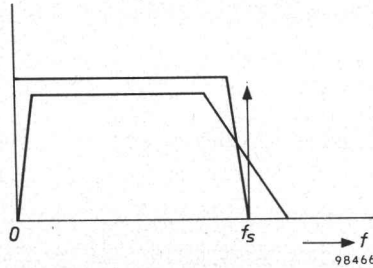


Fig. 25. Subcarrier transmission system for two signals. Both signals have the same bandwidth f_s which is equal to the channel bandwidth. One signal is transmitted in the normal way, the other signal is transmitted as a single-sideband modulated subcarrier with frequency f_s .

If it is required to transmit a signal u_3 in addition to u_1 and u_2 , a second subcarrier modulated by u_3 could be introduced. Another method, however, is to modulate one subcarrier by u_2 as well as u_3 in such a way that they can be separated by synchronous detection. Then the transmission signal becomes

$$u_1 + \sqrt{2} u_2 \cos \omega_s t + \sqrt{2} u_3 \sin \omega_s t \quad (66)$$

Multiplying by $\sqrt{2} \cos \omega_s t$, we get

$$u_2 + \sqrt{2} u_1 \cos \omega_s t + u_2 \cos 2\omega_s t + u_3 \sin 2\omega_s t.$$

After filtering, the result is

$$u_2 + \sqrt{2} u_1 \cos \omega_s t. \quad (67)$$

When the transmission signal is multiplied by $\sqrt{2} \sin \omega_s t$ and the same filtering process employed the result is

$$u_3 + \sqrt{2} u_1 \sin \omega_s t. \quad (68)$$

In this way the three signals can be separated, and are represented by the transmission signal and the two demodulator output signals. However, single-sideband transmission is not possible without introducing cross-talk between u_2 and u_3 .

The method of transmitting three signals as represented by (66) is used in the colour system introduced in the U.S.A. which is usually known as the N.T.S.C. system. This system will be discussed in III.4.

III.3.2.1 Comparison of subcarrier system and dot-interlace system

Now comparison can be made between the subcarrier system and the dot-interlace system, both for duplex transmission. This gives rise to the following observations.

In the dot-interlace system the input signal fed to the transmission channel is

$$\begin{aligned} u_1 h_1(t) + u_2 h_2(t) = \\ u_1 (1 + 2\cos\omega_0 t + 2\cos 2\omega_0 t) + u_2 (1 - 2\cos\omega_0 t + 2\cos 2\omega_0 t) = \\ (u_1 + u_2) + 2(u_1 - u_2) \cos\omega_0 t + 2(u_1 + u_2) \cos 2\omega_0 t, \end{aligned}$$

apart from a proportionality constant. The nominal bandwidth of the transmission channel is f_0 .

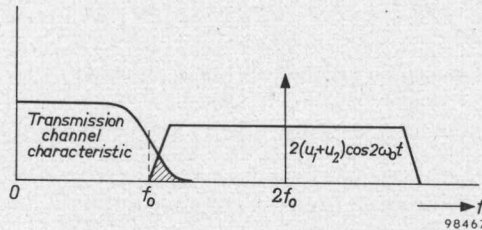


Fig. 26. Dot-interlace system for two signals. The contribution to the transmission-channel output signal of the component $(u_1 + u_2) \cos 2\omega_0 t$ present in the input signal of the transmission channel.

If the signal frequencies of u_1 and u_2 do not extend beyond f_0 , the component $(u_1 + u_2) \cos 2\omega_s t$ will only contribute to the transmission channel output signal in the measure that the transmission channel characteristic extends beyond f_0 . (See hatched area in fig. 26.) This contribution compensates the attenuation of components just below f_0 . This will be clear from III.3.1, where it is pointed out that a symmetrical characteristic may replace an ideal

characteristic, so that the deviation from the ideal form below f_0 can be corrected by a deviation from zero above f_0 . It can also be computed in terms of frequency components.

If the cut-off of the transmission characteristic is not too gradual the term $(u_1 + u_2)\cos 2\omega_0 t$ can be ignored. In that case the received signal can be said to consist of $(u_1 + u_2)$ combined with the subcarrier f_0 single-sideband modulated by the difference signal $u_1 - u_2$.

If the single-sideband modulation is denoted by the operator Ω the transmission signal can be written as

$$(u_1 + u_2) + \Omega (u_1 - u_2). \quad (69)$$

In order to obtain the signal u_1 at the receiver, the received signal is multiplied by $h_1(t)$ in the proper amplitude and phase. If the cut-off of the receiver filter is not too gradual the term $2 \cos 2\omega_0 t$ is of only minor importance, for the situation than arising is similar to that represented in fig. 26. If this term is omitted, therefore, the input signal of the receiver filter will be

$$\{ (u_1 + u_2) + \Omega (u_1 - u_2) \} (1 + 2 \cos \omega_0 t)$$

and the output

$$(u_1 + u_2) + \Omega (u_1 - u_2) + \Omega (u_1 + u_2) + (u_1 - u_2) = 2 (u_1 + \Omega u_1). \quad (70)$$

This analysis in terms of modulated subcarriers shows clearly the correspondence with and the difference from the direct use of subcarriers. In the latter case the transmission signal and the signal reproduced are, see fig. 25,

$$u_1 + \Omega (u_2). \quad (71)$$

In both systems persistence of vision is used to get a correct impression which arises solely from the combination of two successive scans. In the subcarrier system, interference in the desired signal is related to the unwanted signals, but the whole signal information is contained in one normal frame period. In the dot-interlace system, however, the unwanted signals cause no interference with the wanted one, and defects are related to the wanted signal itself and consist of the absence of half the picture elements in one normal frame period.

Before the difference between subcarrier system and dot-interlace system is laid down in a general conclusion we shall examine the receiver circuitry, as its complexity or simplicity largely determines the practical applicability of a system. In this respect our consideration can be restricted to the separation of the television signals received as a composite signal, because this will constitute the main difference from normal receiver circuitry.

III.3.3. Signal separation at the receiver

With both methods of transmitting several signals in a single channel the separation of the signals at the receiver, as far as considered, must be effected by a process known as synchronous detection. The modulated subcarrier system requires the presence in the receiver of a waveform $\cos \omega_s t$ and the dot-interlace system needs:

$$h(t) = (1 + 2 \cos \omega_0 t + 2 \cos 2\omega_0 t).$$

These separating waveforms must be available at the receiver with the right frequency, phase and composition. The information about frequency and phase has to be conveyed to the receiver by an extra synchronizing signal and one example of this is the "burst" introduced in the back porch of each line sync pulse in the American colour-television system. If synchronous detection is used, therefore, the receiver must contain devices for generating these synchronized separating waveforms and for multiplying the received signal by them. This necessarily involves the use of rather complicated circuitry, which must be considered as a disadvantage of this method of signal separation. The possibility for simplifying the separation process will be considered.

III.3.3.1. Non-synchronous detection of a subcarrier

A simplification of the synchronous detection is obtained when the subcarrier is always present with the reference phase in transmission, because it can be filtered out and can act as a synchronizing waveform for the generator. When the transmission signal is

$$u_1 + (u_2 + k) \cos \omega_s t, \quad (72)$$

where u_2 is always positive and varies between 0 and $u_{2\max}$ and where k is a constant >0 , the subcarrier is always present in the right phase. However, this transmission signal has the disadvantage that the peak value of the subcarrier amplitude is much larger than is necessary for obtaining a given signal-to-noise ratio at the receiver. When detection is synchronized the transmission signal can with advantage be given the form

$$u_1 + (u_2 - \frac{1}{2} u_{2\max}) \cos \omega_s t \quad (73)$$

in order to obtain the same signal-to-noise conditions for u_2 and with a subcarrier amplitude $\frac{1}{2} u_{2\max}$ instead of $u_{2\max} + k$.

Of course, an unnecessary increase in the amplitude of the subcarrier is undesirable, especially in a transmission which is intended to be "compatible", in the sense that a normal standard receiver will be able to reproduce the transmission signal as a satisfactory picture. The price of simplifying the receiver circuitry, therefore, is an increase of interference in normal recep-

tion, resulting from the increased subcarrier amplitude. However, simplification is considerable, because the resulting waveform of the transmission signal corresponding to (72) permits the use of non-synchronous detection. In special cases this may make up for decreased compatibility and the worse signal-to-noise conditions for u_2 . Non-synchronous detection of $(u_2 + k) \cos \omega_s t$ in the conventional way (fig. 27a) requires that the bandwidth of u_2 shall be small with respect to f_s . In other cases the circuit of fig. 27b with double-phase detection has to be employed.

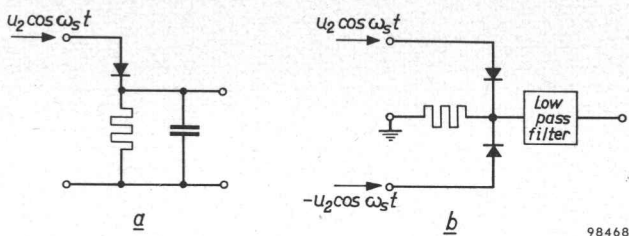


Fig. 27. Conventional detection circuitry. a. Envelope detection. b. Full-wave rectification.

No account has so far been taken of the signal u_1 , which is also present in the same band. As well as detecting the subcarrier, the detection circuit should transform the interfering signal u_1 into signal components that cancel out in two successive scans. This will hold if the detector is able to interpret the signal components of u_1 as sideband frequencies of f_s . It is obvious that this is impossible if, in some interval, the subcarrier amplitude is only small or even entirely absent. For instance, when subcarrier amplitude is zero the output-signal delivered by the detection circuit as derived from u_1 will have the same shape in successive frames (apart from any change in u_1 within one frame period). In order to ensure the presence of a cancelling u_1 -interference in u_2 the subcarrier amplitude must exceed a certain value which depends on the peak amplitude of the u_1 -interference that is present at the input of the detection circuit. The peak value of the modulated subcarrier, already greater than the minimum possible value in order to make conventional detection possible, must therefore be increased once again to ensure the cancellation of the cross-talk of u_1 in u_2 .

As it is important to keep this value as small as possible, another method of avoiding non-cancelling u_1 -interference has been investigated. The special detection method represented in fig. 28 turned out to be an improvement on normal methods. The restriction has to be made that the bandwidth of u_2 is less than $\frac{1}{2} f_s$, so that full-wave detection is not necessary.

The incoming signal is fed to a band-pass filter and its output is detected by the two diodes in fig. 28 in which one follows the upper envelope and the

other the lower envelope. As the upper envelope may extend below, and the lower envelope above the zero-carrier level, owing to the signal u_1 , the diodes have to be biased by $+V_0$ and $-V_0$ respectively. In fig. 29 an example with a special waveform is given. The output signals of the two detectors have the carrier envelopes in opposite phase, but the u_1 interference is partly in the same phase. The signals are therefore subtracted and in fig. 28 this process is carried out in a double triode with common cathode. The resulting output is fed to a low-pass filter of bandwidth u_2 .

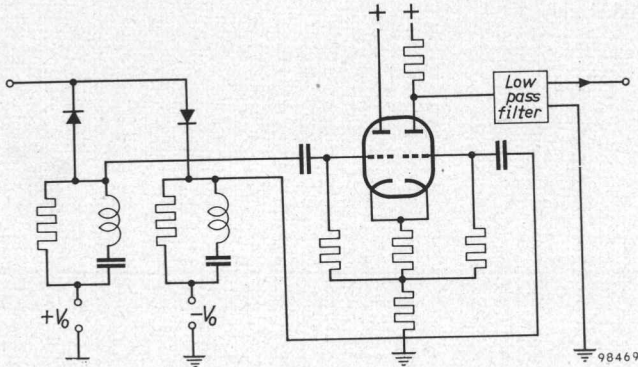


Fig. 28. A double envelope detection circuit. One diode follows the upper envelope, the other the lower envelope. The outputs are subtracted.

Compared with the conventional single detector circuit, this arrangement showed experimentally a considerable reduction of non-cancelling interference. The circuit does not alter the amount of cancelling interference. The proper functioning of the circuit depends on the shape of the interfering

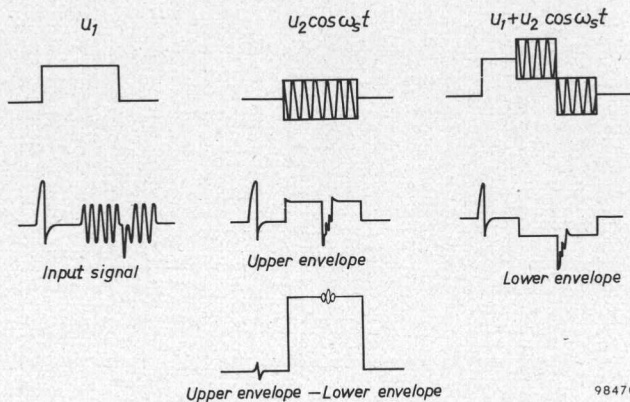


Fig. 29. An example of the double envelope detection process, in the case of a special waveform. The non-cancelling u_1 -interference is considerably reduced due to the subtraction of the envelopes.

u_1 -signal in the detector input. For example when in some interval this signal is a single frequency close to the subcarrier frequency and the subcarrier itself is absent, this u_1 -signal will be detected as if it were a carrier. Under these circumstances the detectors cannot follow the u_1 -signal waveform, nor can they develop equal outputs to neutralize each other by subtraction. A single frequency, however, is seldom present in a television signal, for the waveforms are usually of step-function type, such as those of fig. 29.

In general it can be stated that in order to produce a satisfactory picture from u_2 , the double-detector circuit of fig. 28 needs a relative peak value for the subcarrier which is considerably less than the value necessary for the ordinary detector circuit. A further advantage is that it permits the use of simpler band-pass and low-pass filters, for frequencies normally suppressed by these filters are now neutralized by subtraction. The slope of the cut-off characteristic of the filters can be flatter and it is, in fact, undesirable that the cut-off should be very sharp. It is even possible to omit the filters altogether, but the undesired signals in the output then become very large and the subtraction process is very critical.

Apart from the circuit just described other methods exist whereby crosstalk of u_1 in u_2 can be decreased. They do not involve special circuitry in the receiver, but only a slight modification to the transmission signal. These measures are best suited for a situation such as was assumed for the circuit of fig. 28, namely where u_1 is the main signal and u_2 a secondary signal of smaller bandwidth, and where, if the signal transmitted is reproduced in a normal receiver, it must be possible to obtain a satisfactory screen image from u_1 alone. (Compatibility.)

First we want to point out that, for single-sideband modulation of f_s , conditions for u_2 are best when the lower sideband is transmitted and when f_s is chosen as high as possible. Under these circumstances the interference in u_2 is as small as possible with respect to transmission noise as well as with respect to the presence of u_1 . Furthermore, it is possible to improve the conditions for u_2 by employing pre-emphasis for the sideband of f_s ¹⁵). This is possible because the interference of $u_2 \cos \omega_s t$ in u_1 is mainly caused by the mean subcarrier amplitude and much less by its sidebands. Hence the quality of the image derived from u_1 will hardly be affected by such a pre-emphasis. The pre-emphasis delivers a better signal-to-cross-talk ratio for the frequency components of u_2 in question. Of course, the receiver has to embody the appropriate arrangements for de-emphasis.

With advantage a pre- and de-emphasis process can be used which shows symmetry with respect to f_s , though only one sideband may be present, as this can be effected with simple circuitry. Apart from simplicity, there is the advantage that the de-emphasizing circuit at the receiver contributes to the necessary band-pass action which is needed for separating the modulated subcarrier from the combined signal.

The above system of pre-emphasis naturally only improves the conditions for the high frequencies of u_2 . For the lower frequencies of u_2 , i.e. the composite signal frequencies near the subcarrier frequency, an improvement is possible in another way¹⁵). This method is based on the addition to the transmission signal of extra signal components derived from u_1 . These components are related to u_1 in such a way that the cross-talk in u_2 due to the presence of u_1 is largely compensated. The desired balance is obtained in the following way.

For each component a $\cos(\omega t + \varphi_1)$ occurring in u_1 , a signal component

$$-a \cos \{(\omega_s + \omega_s - \omega) t + \psi\}$$

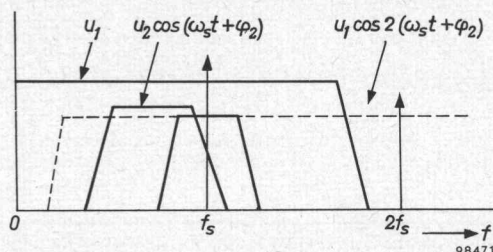


Fig. 30. Compensation of cross-talk. A subcarrier is modulated by the signal u_2 . Cross-talk in this signal from the main signal u_1 is compensated by the addition of the lower sideband of $-u_1 \cos(2\omega_s t + 2\varphi_2)$ near the subcarrier frequency.

is introduced into the composite signal in a phase such that in detection of the subcarrier both components gave rise to cross-talk components of the same frequency and phase but of opposite sign (fig. 30). It can also be said that the signal components of u_1 and the added signal components both represent a single-sideband amplitude modulation of the subcarrier which modulations are of opposite polarity. It can be easily shown that, when the subcarrier is given by

$$\cos(\omega_s t + \varphi_2),$$

the compensating condition is obtained if

$$\psi = 2\varphi_2 - \varphi_1.$$

Therefore the additional signal is

$$-u_1 \cos(2\omega_s t + 2\varphi_2). \quad (74)$$

Of course, only that part of this signal is taken into account which is near the subcarrier frequency. (See fig. 30.)

This extra contribution to the transmission signal will cause a visual effect in the reproduction of u_1 which is only of minor importance. In the first place it is correlated to u_1 , in the second place it covers only a very limited frequency

range and in the third place the visual effect can be further diminished by a suitable choice of subcarrier phase properties. Figure 31 shows an arrangement for compensating cross-talk in this manner. In the modulator M_1 the subcarrier is modulated by u_2 . The modulated subcarrier is added to the signal u_1 .

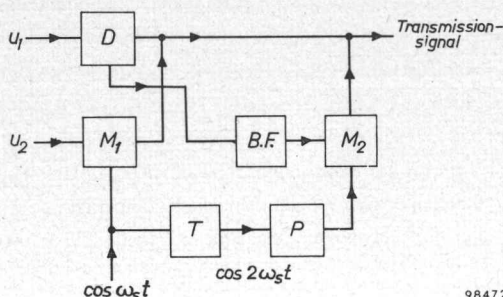


Fig. 31. Block diagram for cross-talk compensation according to fig. 30. M_1 = subcarrier modulator, B.F. = band-pass filter, M_2 = modulator, T = frequency doubler, P = phase shifter, D = delay line.

From u_1 a narrow band of frequencies around f_s is filtered out by the band filter $B.F.$ This signal is modulated in M_2 on the double subcarrier frequency obtained in the stage T , adjusted to the appropriate phase by stage P . The delay which occurs in $B.F.$ is compensated by a corresponding delay conferred on u_1 in D .

III.3.3.2. Non-synchronous detection in a dot-interlace system

The possibility of using non-synchronous detection for the separation of the signals of a two-signal dot-interlace system will now be considered. According to (69) in III.3.2.1 the transmission signal can be written in the simplified form

$$u_1 + u_2 + \Omega (u_1 - u_2).$$

It will be clear that the carrier with single sideband present in this signal cannot be detected by conventional envelope detection, as $u_1 - u_2$ will show negative values. The addition of subcarrier is necessary. The double-detection circuit of fig. 28 cannot be used here, because it can only detect a signal with a relatively small bandwidth. With dot-interlace the bandwidths of the direct signal ($u_1 + u_2$) and the modulation ($u_1 - u_2$) are in principle the same.

With the subcarrier added the transmission signal becomes

$$u_1 + u_2 + \Omega (u_1 - u_2 + u_{2\max} + k) \quad (75)$$

where k has the same significance as in (72). In the video stage the signal can now be detected by a full-wave rectifier. In the intermediate frequency stages of a receiver the transmission signal will appear as modulation on a vision carrier of frequency f_p . Two carriers, f_p and $f_p + f_0$, are therefore present, modulated

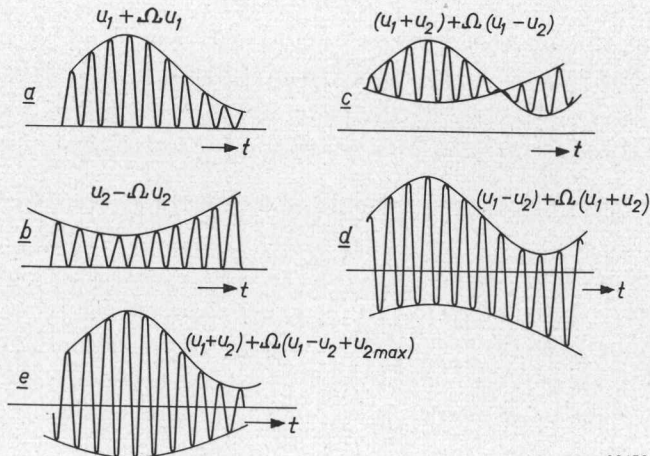
by $u_1 + u_2$ and $u_1 - u_2$ respectively. Each carrier can be detected after the other has been suppressed and each detected output signal, $u_1 + u_2$ and $u_1 - u_2$, suffers interference from the sidebands of the unwanted carrier. The information corresponding to u_1 and u_2 can be obtained from the sum and the difference of the two detector output signals but, in consequence of the suppression of the undesired carrier, the result will show some attenuation of the highest frequencies of u_1 and u_2 .

When detection is carried out in the video stage it is not advisable to demodulate the subcarrier by a full-wave detector because the signals can be separated more simply. In order to explain the principle of this method of separation, it is necessary to consider the shape of the signals occurring in a two-signal dot-interlace system. The output signal of the transmission channel for the input

$$u_1 h_1(t)$$

will have the form shown in fig. 32a. From the analysis of the dot-interlace system it follows that for $\omega_0 t = 2n\pi$ this signal equals u_1 and for $\omega_0 t = \pi + 2n\pi$ this signal is zero. When the slope in u_1 is not too large, negative excursions will be very small.

It will be clear that for u_2 the result is as illustrated in fig. 32b. As can be seen, the zeros of fig. 32b coincide with the u_1 values and the zeros of fig. 32a coincide with the u_2 values. The transmission signal has the form of fig. 32c.



98473

Fig. 32. Waveforms occurring in a two-signal dot-interlace system.

- a) Output signal $u_1 + \Omega u_1$ of the transmission channel for the input $u_1 h_1(t)$.
- b) Output signal $u_2 + \Omega u_2$ for input $u_2 h_2(t)$.
- c) Output signal for input $u_1 h_1(t) + u_2 h_2(t)$.
- d) Output signal for input $u_1 h_1(t) - u_2 h_2(t)$.
- e) Output signal for input $u_1 h_1(t) + u_2 h_2(t) + 2u_{2max} \cos \omega_0 t$.

When, however, the signals of fig. 32*a* and 32*b* are combined in opposite polarity the transmission signal is modified, and is like that in fig. 32*d*. Above the zero level the signal contains mainly information about u_1 , and below the zero level mainly information about u_2 . For the signal composed according to (75) the form is similar, but the envelope u_2 is of opposite sign. (See fig. 32*e*.) At the receiver these signals, fig. 32*d* and fig. 32*e*, can be separated into two parts by means of a simple clipping circuit. The outputs differ only slightly from the signals represented in fig. 32*a* and fig. 32*b*. However, they are not exactly identical and a certain amount of cross-talk will be present. Measures may be taken to reduce this effect. This can be done by pre-correcting u_1 and u_2 ^{16,17}).

Comparison of fig. 32*d* and fig. 32*c* shows again that the benefit of the non-synchronous operation must be paid for in an increase in the peak value, i.e. in worse signal-to-noise conditions.

III.3.4. Conclusion

The discussion about the use of subcarriers and dot-interlace may be summed up as follows.

In a subcarrier system signals u_1 , u_2 , u_3 etc. are all transmitted in different ways. Signal u_1 is transmitted normally, but the others are modulated on subcarriers of different frequencies, a single subcarrier being modulated by not more than two of these signals. Because of this, the subcarrier system is particularly suited to the transmission of signals of dissimilar qualities; e.g. signals which differ in bandwidth or peak value.

In the dot-interlace system, on the other hand, signals u_1 , u_2 , u_3 etc. are all transmitted in the same way. Before transmission they are all transformed by a sampling process which differs for the different signals only in relative phase. The dot-interlace system is, therefore, best suited to the transmission of signals having similar characteristics; e.g. signals which have the same bandwidths and peak values.

With both systems the reconstituted signals at the receiver contain an interference effect which cancels out in two successive scans. In the subcarrier system this interference in a certain desired signal u_k is related to the undesired signals. As a result, when for some reason cancellation is not complete, there will be cross-talk of the unwanted signal in the desired one. However, when the unwanted signal remains zero for some period no interference at all will be experienced. The subcarrier system is, therefore, a particularly suitable one when it is desired to keep one of the signals as free from interference as possible. In colour television, for example, "compatibility" requires that there should be a minimum of interference with the normal transmitted signal. Moreover the ultimate result in colour television is not very sensitive to mutual cross-talk

because there is a large degree of correlation between the three signals which combine to produce the coloured image.

In the dot-interlace system interference with the separated signals at the receiver is related only to the desired signals themselves. Cross-talk is therefore absent, in principle, but the interference is always present in the same magnitude in all signals. Because of this the system is most suited to cases where the signals are very sensitive to cross-talk, when for example they are entirely uncorrelated.

So far the most important application of multiplex transmission of television signals is to colour television. When compatibility is needed it appears that the subcarrier system is the better, of the reasons being:

- (1) The colour-picture information is most advantageously transmitted by three signals of which one has the full channel bandwidth and the other two much smaller bandwidths. (See II.3.4.)
- (2) The alternating signal components in the transmission signal have to be as small as possible and their frequencies as high as possible.
- (3) The correlation which exists between the three signals reduces the sensitivity to mutual cross-talk.

In III.4 two forms of a colour television system using subcarriers will be described.

III.3.5. Experiments

Having completed our theoretical study of dot-interlace and subcarrier systems, we shall now give a brief account of our experiments in connection with these transmission systems. As far as dot-interlace is concerned, we built a dot-interlace system for transmitting one signal, and another for transmitting two signals. The original video signals had a bandwidth of 5 Mc/s. The dot

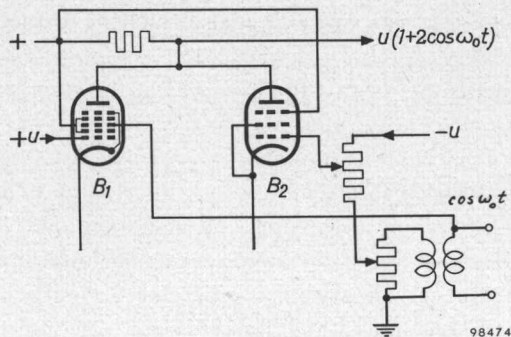


Fig. 33. Sampling circuit for dot-interlace system. In the hexode B_1 the video signal u is modulated on $\cos \omega_0 t$. By the addition of u and $\cos \omega_0 t$, in opposite polarity, the desired from $u(1 + 2 \cos \omega_0 t)$ is achieved.

frequency f_0 was therefore 5 Mc/s. Sampling at transmitter and receiver with a frequency of 5 Mc/s was effected by the operation

$$u(1 + 2 \cos \omega_0 t).$$

Figure 33 shows the principle of the circuit employed. In the hexode B_1 the input signal u is modulated on $\cos \omega_0 t$. The output signal of a single modulator like this will have the form

$$(u + k_1)(1 + k_2 \cos \omega_0 t).$$

By the addition of u and $\cos \omega_0 t$ in opposite polarity and with proper amplitude by B_2 the desired form $u(1 + 2 \cos \omega_0 t)$ can be achieved. The decoder at the receiver was quite similar.

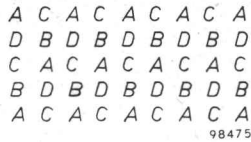


Fig. 34. Dot pattern for a sampling frequency which is an odd multiple of half the line frequency.

The sampling frequency was made an odd multiple of half the line frequency. As stated in II.3.1, the phase alternating in successive scans of one line is ensured; the dot pattern corresponds to fig. 34. (See II.3.2.2 fig. 19a.) The frequency relationship between sampling- and line frequency was obtained by means of an arrangement like that in fig. 35. The sine-wave frequency of the oscillator O is brought to half the line frequency by the divider D which contains binary counters and feedback features as shown in fig. 17 of part II. The output signal has a rectangular shape. It is compared to pulses of half the line frequency which are provided by subjecting pulses from the line time-base, $T.B.$, to frequency division, the discriminator P embodying a bistable multivibrator which operates in a way similar to the phase discriminator described in II.3.2.1, fig. 15. The d.c. component present in the output of the multivibrator is fed to a reactance tube, which controls the oscillator frequency in such a way that the desired frequency relation is maintained.

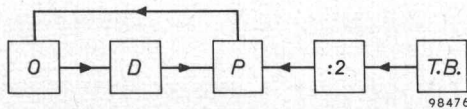


Fig. 35. Arrangement for obtaining the relationship between sampling and line frequency
 O = Oscillator, D = frequency divider with binary counter, P = phase discriminator
 $T.B.$ = time base.

In the dot-interlace system with one signal the channel bandwidth was $\frac{1}{2}f_0 = 2.5$ Mc/s. The transmission filter was built up from conventional m -derived π -sections without any special precautions being taken to fulfil the special filter requirements. It was found that very good results could nevertheless be obtained. The results are shown in fig. 36, 37 and 38 on p. 121 and 122. Normal transmission is represented by fig. 36. Normal transmission with half the bandwidth is represented by fig. 37. The dot-interlace system is represented by fig. 38. It can be seen that picture resolution in fig. 38 is far better than in fig. 37 and approximates the picture resolution of fig. 36.

In the dot-interlace system for two signals the transmission characteristic appeared to be much more critical. This could have been expected because any shortcoming in the characteristic will show up as a cross-talk of the undesired signal in the desired signal. The pictures were totally different so any cross-talk is perceived very rapidly.

In employing synchronous detection it was found that errors due to shortcomings in the transmission characteristic could be compensated to a large extent by a certain misadjustment of the receiver sampling. The sampling function, being normally

$$\{1 \pm 2 \cos (\omega_0 t + \phi)\}$$

where ϕ represents the phase shift for frequency f_0 in the transmission path, was converted into

$$1 \pm a \cos (\omega_0 t + \phi + \delta\phi)$$

where a and $\delta\phi$ were adjusted experimentally for optimum result. Apparently the sampling instants for zero, or at least minimum, response of the undesired signal did not coincide with the minima and maxima in the waveforms $u_1 - \Omega u_1$ and $u_2 + \Omega u_2$ as represented in fig. 32. In other words, the envelope of Ωu shows a certain deviation from u , and a better correspondence with u could be found for sampling instants slightly displaced from the maxima of Ωu . The distortion of the envelope is, somewhat exaggerated, shown in fig. 39 for a rectangular waveform when minimum-phase filters are employed.

Non-synchronous detection was also investigated, a transmission signal was used of the kind shown in fig. 32c. In that case of course the compensation of faulty filter-characteristic influence cannot be effected by a phase shift in receiver sampling. The deviations due to the filter shortcomings, mainly in its phase characteristic, are shown in fig. 39¹⁸). It can be seen that for the rectangular wave form the lower envelope is certainly not without modulation. In this case correction was obtained by pre-correcting u_1 and u_2 . With this pre-correction an acceptable picture could be obtained. The separation was effected by a mere clipping action with simple diode circuitry. Far better results were obtained, however, by the employment of phase equalization networks

and by the use of filters with bridged T -structure in the transmission path. In these cases the pre-correction could be considerably decreased. No attempt was made to build a filter with a characteristic conforming to the principles described in III.3.1.1.

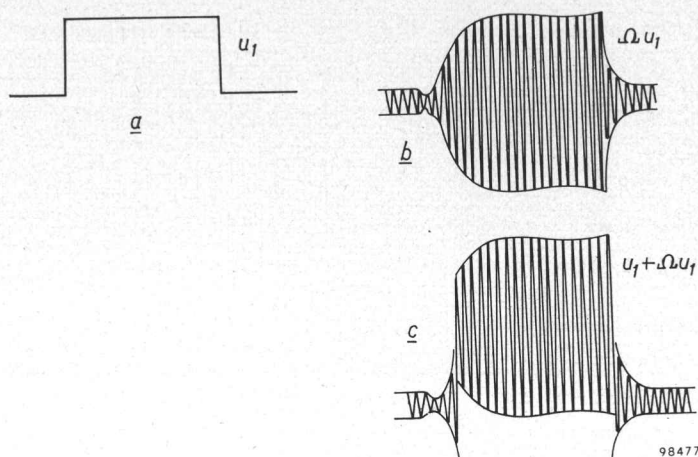


Fig. 39. The shape of u_1 , Δu_1 and $u_1 + \Delta u_1$ for a rectangular waveform and a minimum-phase transmission characteristic (somewhat exaggerated).

The experiments with sub-carriers were mainly carried out in connection with colour-television transmission. These experiments are described in III.4.2. For duplex transmission experiments were done with the transmission signal

$$u_1 + u_2 \cos \omega_s t$$

in the form of fig. 25. As could be expected, we found in reproducing this transmission signal that cross-talk of u_2 occurred owing to the non-linearity of the picture tube, which demodulates $u_2 \cos \omega_s t$ to some extent. By adding u_2 in proper amplitude and polarity to the transmission signal the cross-talk could be compensated and reduced below the level of visibility.

III.4. The transmission problem in colour television

The possibilities described in section III.3 for a multiplex transmission within a frequency band which is considerably narrower than the added bandwidths of the separate signals have been exploited in a number of colour-television transmission systems. In this section a description is given of these systems, special attention being paid to two of them, the transmission system now in use in U.S.A., the so-called N.T.S.C. system, and a system which has been developed by ourselves, and which we call the two-subcarrier system.

Many of the basic problems and principles of colour television have already

been considered in section II.3.4. Summarizing this we can give the following description of the colour-television transmission problem: If an existing black-and-white television service is to be developed into a colour-television service, without making the black-and-white receivers in the hands of the public obsolete, the transmission system has to be "compatible". This requirement of compatibility means that the information of a colour picture must be transmitted within a normal television channel and in such a way that existing black-and-white receivers can reproduce the transmitted picture in monochrome, without any modification to the receiver *). The consequences of the compatibility requirement are that the three signals necessary to determine luminance and colour must be transmitted simultaneously and that one signal has preferably to be made similar to the signal in black-and-white television. This signal — the luminance signal — is transmitted in the normal way with normal bandwidth. The other signals have to be added to this main signal, within its frequency-band, in such a way that in black-and-white reception their presence is hardly perceptible. To some extent this is possible by exploiting the persistence of vision as it is done in dot-interlace and sub-carrier systems described in section III.3. These added signals can have relatively small bandwidths, because the eye is relatively insensitive to errors in the colour of picture detail. When the right luminance is present in the detail, such errors are hardly noticeable. In the N.T.S.C. system and the two-subcarrier system the addition of the second and third signal to the main signal — the luminance signal — is effected by using the subcarrier principle. This is in accordance with our conclusion, given in III.3.4, that the subcarrier principle is best suited for colour television transmission.

The main difference between the two systems mentioned is in the number of subcarriers and the method of decoding at the receiver. In the N.T.S.C. system the colour signals are modulated on two subcarriers of the same frequency but differing in phase; in other words only one subcarrier is employed. In the receiver synchronized detection is employed in order to recover the three video signals. In the two-subcarrier system the subcarriers have different frequencies and signal separation at the receiver is non-synchronous. We shall first give attention to the N.T.S.C. system.

*) In the sense in which it is defined here, compatibility is not a generally accepted requirement of a colour-television system. It is in U.S.A., but elsewhere it has been suggested that compatibility requires only that a monochrome receiver shall, without alteration be able to reproduce a monochrome picture from a colour transmission. In this case, all the information content of a colour transmission need not lie within the band of the monochrome receiver, but only that part of the total information which is necessary for the reproduction of a monochrome picture. Such a system might be called one of reduced compatibility and the principles of both systems described in this chapter can be used in it as in a fully compatible system.

III.4.1. The N.T.S.C. system

The N.T.S.C. system has been developed in U.S.A. as a combined effort by several firms. It was officially introduced in U.S.A. at the end of 1953, and probably this system or a system very similar to it will be employed in other countries. A detailed description of this system can be found in the literature cited in the references ^{19,20,21,22}). Here only a brief description will be given.

As stated in the introduction, the N.T.S.C. system is a subcarrier system in which one subcarrier is employed, being modulated in two different phases and added to the main signal, the luminance signal. The modulating signals consist of combined colour-difference signals, the latter being signals proportional to the difference between the luminance signal and one of the primary signals. The modulating signals are of different bandwidths. The wideband signal (1.5 Mc/s) is called the *I* signal. It refers to the position of the original colour in the colour triangle with respect to the complementary colours orange and cyan. The narrow band signal or *Q* signal (0.5 Mc/s) refers to the location with respect to the complementary colours green and magenta. Modulation takes place in such a way that the subcarrier amplitude is zero where no colour is present in the picture.

Owing to the fact that the subcarrier is modulated in two phases, differing by ninety degrees, the result finally obtained by this modulation can also be interpreted as a single subcarrier modulated in amplitude as well as in phase. In virtue of the choice of modulating signals, the subcarrier phase refers to the hue of the colour and the subcarrier amplitude refers to the saturation of the colour.

After modulation, the resulting modulated subcarrier is added to the luminance signal, which thereby acquires unequal sidebands. The lower sideband has the bandwidth of the *I* signal. The upper sideband is equal to the *Q* signal. The composite video signal of the N.T.S.C. system is shown in fig. 40.

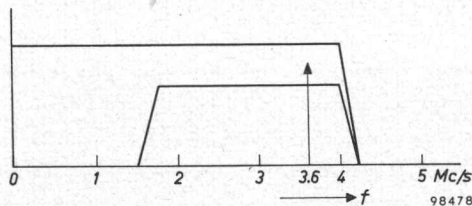


Fig. 40. Frequency spectrum of the composite video signal of the N.T.S.C. system.

Decoding in the receiver of the received composite signal is effected by synchronous detection of the subcarrier delivering the colour-difference signals. By proper matrixing of the luminance signal and the colour-difference signals the three primary signals are recovered and are fed to the display. On the recei-

ver screen, picture detail is reproduced correctly in regard to brightness. The reproduction of the colour detail is matched to the characteristics of the eye with respect to colour perception and to the statistics of colour transients as described in section II.3.4. In order to make synchronous detection at the receiver possible, a reference frequency and a reference phase of the subcarrier have to be known at the receiver. The required information is conveyed in what is called a "burst" giving a number of periods of the unmodulated subcarrier in a reference phase during the back porch of the line synchronization wave form.

It will be clear that the N.T.S.C. system is a compatible system. Apart from the compatibility requirement however, the transmission of the luminance signal as main signal is advantageous in colour perception. The luminance of the colour picture in this case is fully determined by the luminance signal and in principle will not be disturbed by shortcomings in the transmission of colour information *)^{23,24}). Now that there is several years experience in colour-television transmission, it can be said that the compatibility requirement did not exclude very attractive possibilities for the colour-television transmission. Even if compatibility is not required, the compatible techniques can be considered as very suitable solutions to the transmission problem.

III.4.2. *The two-subcarrier system*

The second system to be considered here is a transmission system which has been developed by ourselves and into which extensive investigations have been carried out^{15,25}). This too is a compatible system based on the use of two subcarriers combined with a luminance signal. In the international deliberations about colour television standards it has for some years been a competitor of the American N.T.S.C. system²⁶).

The two subcarriers employed in this system were of different frequencies and were modulated with the red and blue primary signal. Figure 41 is a block diagram. Starting at the scene to be transmitted in colour, we see that the first component of the system is a camera suitable for generating three primary signals corresponding to the green, red and blue light components of the scene. These signals are indicated by E_G , E_R and E_B . They are assumed to vary between 0 and 1. These primary signals are fed to a matrix unit which delivers at its output the luminance signal

$$E_Y = \alpha_0 E_G + \beta_0 E_R + \gamma_0 E_B. \quad (77)$$

The coefficients α_0 , β_0 and γ_0 correspond to the sensitivity of the human eye

*) This is true of a linear system. Owing to non-linearities in the system — such as the picture tube characteristic and the inherent gamma correction measures at the transmitter — the transmission of colour information will have a certain influence on that of luminance.

to green, red and blue light respectively. They are given values such that

$$\alpha_0 + \beta_0 + \gamma_0 = 1. \quad (78)$$

Hence the luminance signal will also vary between 0 and 1.

The luminance signal is transmitted in the normal way with the normal bandwidth of 5 Mc/s (625-lines standard). In addition, the red and blue signals are transmitted with the aid of two subcarriers each being modulated by one of these signals. As stated in II.3.4 it is possible to reduce the bandwidths of these signals, in virtue of the lower sensitivity of the eye to colour contrast in picture details. In our experiments the bandwidth was reduced to about 2 Mc/s for the red signal and 1 Mc/s for the blue signal. After filtering, the red and blue signals modulate the subcarriers, which fall within the normal video range (0—5.5 Mc/s) and have a special frequency and phase relation with the scanning frequencies (to be discussed later). Their frequencies are made as high as possible; however, a frequency difference of about 1 Mc/s must exist between the

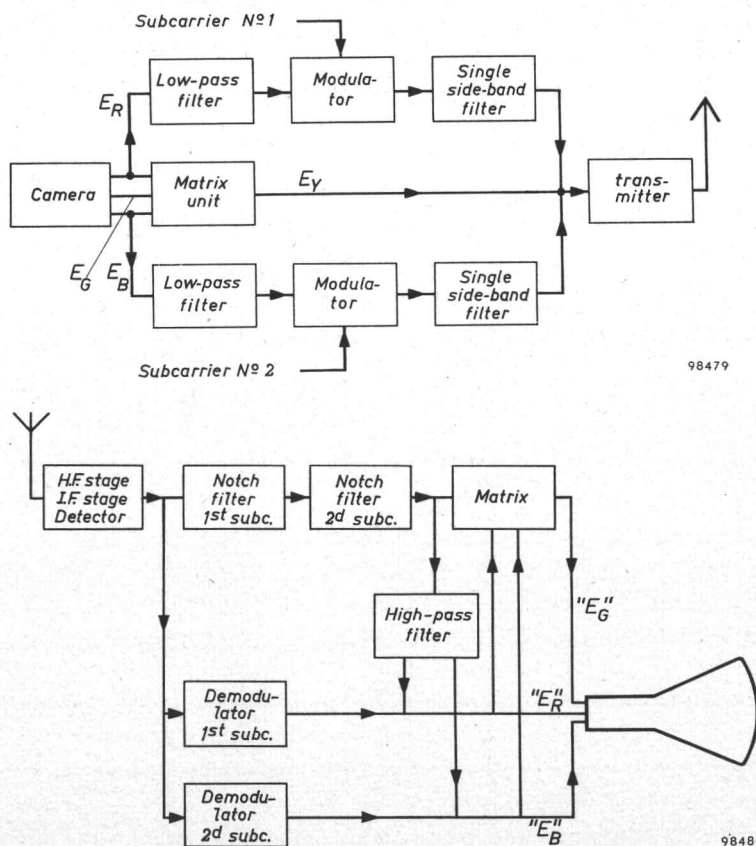


Fig. 41. Block diagram of the two-subcarrier system. a) Transmitter, b) Receiver.

higher subcarrier frequency and the sound carrier at 5.5 Mc/s, in order to make it possible for the subcarrier to be separated with a practical receiver filter. The subcarrier frequencies were about 3.6 Mc/s ($\frac{1}{2} \times 3 \times 3 \times 3 \times 17 \times f_1$) and 4.6 Mc/s ($3 \times 3 \times 3 \times 11 \times f_1$) where f_1 is the line frequency. See fig. 42.

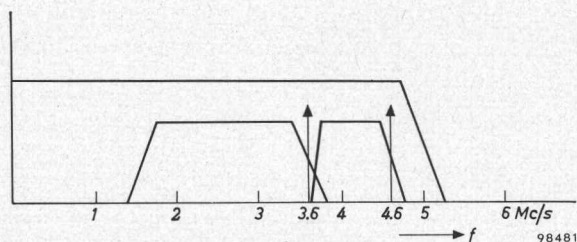


Fig. 42. Frequency spectrum of the composite video signal of the two-subcarrier system.

After modulation the subcarriers are added to the luminance signal. In principle, this modulation may be double-, vestigial, or single-sideband modulation. The signal-to-noise conditions of the colour signals at the receiver, however, are optimum with single-sideband modulation, provided the signal amplitude is the same in the three cases. The composite signal is fed to a transmitter and modulated on a picture carrier.

Up to the first video amplifier, the receiver is practically identical to a normal black-and-white receiver. The second detector delivers a signal which is assumed to be the same as the composite video signal present at the transmitter. This signal is fed to two demodulators, each consisting of a band-pass filter tuned to the appropriate subcarrier and a detector circuit. The observations about detection made in III.3.3.1 apply to these detector circuits, and the principle of fig. 28 is employed in them.

The red and blue signals with small bandwidth are recovered at the outputs of the demodulators. In order to increase the sharpness of the images reproduced by these signals, the high frequencies of the luminance signal (mixed highs) are added with proper amplitude and time delay. These frequencies are obtained by means of simple high-pass filters. An equivalent green signal is formed in a matrix unit where demodulator output signals are combined with mixed highs from the luminance signal. The three signals thus obtained are fed to a display device suitable for reproducing a colour picture. This device may contain three projection tubes and dichroic mirrors, or may be some type of direct viewing colour tube^{27,28,29}). When the system is well adjusted, large areas in the picture will appear in the original colour, whereas the high video frequencies are reproduced in black and white.

It can be stated that an important advantage of this system is the use of non-synchronous detection, which allows subcarrier generating and synchronizing

circuits as well as synchronized detectors to be omitted from the receiver. Moreover, a synchronizing signal for the subcarrier like the burst in the N.T.S.C. signal is not needed. However, such a burst may be useful for other purposes. As only the subcarrier amplitudes carry the colour information, the right amplitude ratio between luminance signal and subcarriers must be maintained very accurately and this ratio must not be affected by different propagation for various parts of the frequency band, changes in amplifiers or mistuning of the receiver. A suitable means of maintaining a correct amplitude ratio in the receiver is to ensure that a fixed amount of the two subcarriers is present at certain predetermined instants (as the burst is); automatic ratio control is than possible¹⁵⁾.

The system has now been briefly explained. The basic idea (viz., the use of two subcarriers within the band of a normal video signal and modulated by primary signals) is not new³⁰⁾. But the inherent problems of such a system have never been studied by others.

Choice of subcarriers

One of the major problems to be solved in the utilization of subcarriers is to give them frequency and phase such that their presence can be tolerated in black-and-white reception. Where one subcarrier is used, all the reasoning and results in section II.3.2.2 apply. However, when two subcarriers are added the situation is much more complicated because the difference frequency of the two subcarrier frequencies has also to be taken into account. This frequency will always be more or less visible in the reproduced picture, owing to the non-linear characteristic of the cathode-ray tube. When for instance two subcarriers are used whose frequencies are synchronized with the line frequency and are equal to odd multiples of half the line frequency (and in these circumstances each subcarrier alone will be scarcely visible, because opposite polarity in successive scans is assured) the difference frequency will be a multiple of the line frequency and very much visible.

The visual effect of the difference frequency has been found to be at least as important as the perceptibility of the subcarriers themselves. This was discovered during experiments carried out to determine the most satisfying frequency and phase properties for the subcarriers. Several possibilities were investigated. Subcarrier frequencies equal to n , $n + \frac{1}{2}$, $n + \frac{1}{3}$ or $n + \frac{1}{4}$ times the line frequency were employed, and phase resetting at the beginning of each field or frame period (involving phase shifts of $\pm 90^\circ$ or $\pm 180^\circ$) was introduced. Each pair of subcarriers was judged by three disturbing effects: the perceptibility of a dot structure, the stroboscopic effect appearing as moving dot structures and the moiré effects caused by interference of dot structure and picture detail.

Finally one combination of two subcarriers was found which presented a satisfactory solution and this was used in all further experiments with the

transmission system. The lower subcarrier frequency is made equal to an odd multiple of half the line frequency. In addition, the subcarrier is given a sudden 90° phase shift, alternately positive and negative at the beginning of each field. The higher subcarrier frequency is made equal to a multiple of the line frequency and this subcarrier is given a phase shift of $+$ and -180° at the beginning of each field.

The first subcarrier can be written as

$$A_1 \cos \left\{ (n + \frac{1}{2}) 2\pi f_1 t \pm \pi/4 + \phi_1 \right\} \quad (79)$$

the positive and negative signs being operative in alternate fields, f_1 is the line frequency and n is an integer (equal to 229). The second subcarrier can be written

$$A_2 \cos \left\{ m 2\pi f_1 t \pm \pi/2 + \phi_2 \right\} \quad (80)$$

in which m is an integer (equal to 297.) In consequence the frequency difference can be represented by

$$\begin{aligned} A_3 \cos \left\{ m 2\pi f_1 t \pm (\pi/2) + \phi_2 - (n + \frac{1}{2}) 2\pi f_1 t \mp (\pi/4) - \phi_1 \right\} = \\ A_3 \cos \left\{ (q + \frac{1}{2}) 2\pi f_1 t \pm (\pi/4) + \phi_3 \right\} \end{aligned} \quad (81)$$

in which $q = m - n - 1$.

This expression is of the same nature as the expression for the first subcarrier. The pattern in the picture corresponding to this has very suitable properties in regard to perceptibility. This has been described in section II.3.2.2; the pattern was shown there in fig. 19d and is given here in fig. 43a. Looking for possible stroboscopic effects, we find that there is no direction in which the dots A , B , C and D appear successively along a straight line. Moreover, the existing stroboscopic effects caused by the succession of A and C , and B and D in both horizontal and vertical directions are all equivalent, so that there exists no direction or preference for the perceptibility of stroboscopic effects. Moving dot structures are therefore less perceptible than those occurring in the pattern of an odd multiple of half the line frequency without sudden phase shifts (fig. 19a in section II.3.2.2) where, in particular, the upward sequence A , B , C and D gives rise to a rather striking stroboscopic effect.

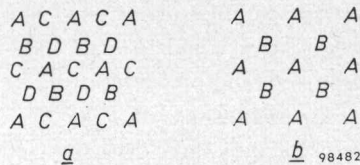


Fig. 43. Subcarrier patterns.

- The lowest subcarrier. The frequency is an odd multiple of the line frequency; 90° phase shifts are given, alternately positive and negative, at the beginning of each field.
- The highest subcarrier. The frequency is a multiple of the line frequency; a phase shift of $+$ and -180° is given at the beginning of each field.

The pattern of the second subcarrier appears in fig. 43b. In this pattern, the dots of the odd fields (*A* and *C*) coincide and so do the dots of the even fields (*B* and *D*). Hence no cancellation effect of two successive scans of the same line is present here. Yet the pattern of the second subcarrier of fig. 43b gave rise to very little disturbance, owing to the high value of the second-subcarrier frequency and the interlacing of the dots of the two fields. (*A* and *B*.)

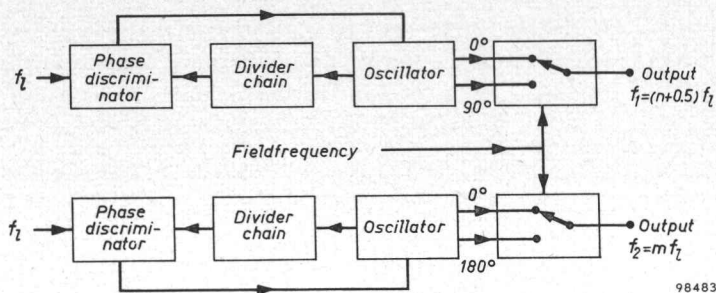


Fig. 44. Subcarrier generators. The oscillator frequency is locked to the line frequency. The oscillator delivers two sine waves differing only in phase. The output is switched electronically at the beginning of each field.

Figure 44 illustrates the way in which the subcarriers were produced. In both generators there is an oscillator locked to the line frequency. The oscillator frequencies are $(n + \frac{1}{2})f_1$ and $m f_1$. The line frequency lock is effected by means of a divider chain and a phase discriminator. The output voltage of the phase discriminator depends upon the phase relation between the two frequencies and controls the oscillator frequency. Each oscillator delivers two sine waves, only differing in phase: 90° for the lower and 180° for the higher subcarrier. The generator output is connected to the oscillator output by means of an electronic switch, which is switched at the beginning of each field period. In this way the desired frequency and phase relations of the subcarriers can be obtained.

Subcarrier amplitudes

In deciding upon the relative subcarrier amplitudes in the composite signal a compromise must be made between compatibility and colour-picture quality. If the amplitude is increased, the annoyance caused by the subcarrier in monochrome reception will increase, and if it is decreased signal-to-noise conditions become worse. In the first experiments the relative subcarrier amplitudes were fixed at 40% and 25%; the composite signal could therefore be written

$$E_Y + 0.4 E_R \cos 2\pi f_1 t + 0.25 E_B \cos 2\pi f_2 t \quad (82)$$

where f_1 and f_2 are the subcarrier frequencies.

The second subcarrier level was made considerably lower than the first-sub-

carrier level because the blue component in the reproduced colour picture is less sensitive to interference. It thus became possible for the signal-to-noise ratio of the blue signal to be lower.

The subcarriers were modulated to a depth of 100% so that subcarrier amplitude was zero in dark areas. In order to avoid cross-talk of the luminance-signal components in the colour signal a special detection circuit — designed on the principles described in III.3.3.1 (fig. 28) — had to be incorporated in the receiver. It was found that only the first subcarrier had to be detected thus. Owing to the lower interference sensitivity of the blue signal and the weaker luminance components in the region of the second subcarrier, it was possible to employ normal envelope detection for the latter.

In later experiments we were able to cut down the subcarrier amplitudes still further by pre-emphasizing the subcarrier sidebands in the manner described in section III.3.3.1. The peak amplitudes of the subcarriers were finally reduced to 25% and 15%. Hence the composite signal was

$$E_Y + 0.25 E_R \cos 2\pi f_1 t + 0.15 E_B \cos 2\pi f_2 t. \quad (83)$$

For these values the peak level of the composite signal is 1.4 (during peak white, when $E_G = E_R = E_B = E_Y = 1$.) Hence the maximum overshoot above white level is 40%. No undershoot below the black level can occur because for the minimum value of E_Y and maximum values of E_R and E_B , i.e. when $E_G = 0$, $E_R = 1$, $E_B = 1$, the luminance signal E_Y is 0.4 and the amplitude of the added subcarriers is also 0.4. The signal range of the composite signal is illustrated in fig. 45a. For the purpose of comparison the levels of the N.T.S.C. composite signal are also shown in fig. 45b. Here maximum excursions may occur of 33% above white and 33% below black level. Though only a difference of 7% exists between the two cases, it cannot be denied that at this point the two-subcarrier system is at a disadvantage compared with the N.T.S.C. system because maximum overshoot occurs more frequently in it. This effect was noticed in practice.

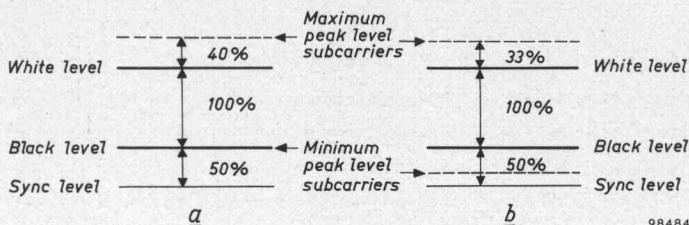


Fig. 45. The signal range of the composite signal: a) for the two-subcarrier system, b) for the N.T.S.C. system.

In the N.T.S.C. system large subcarrier amplitudes occur when the colour has a high degree of saturation. In the two-subcarrier system maximum overshoot occurs in white.

The stringent requirements for the two-subcarrier system on the amplitude balance of the luminance signal and the subcarriers, as outlined before, appeared to result in less difference from the N.T.S.C. system than was expected, for automatic gain control could be obtained fairly simply at both transmitter and receiver. For this purpose, reference signals like the colour burst in the N.T.S.C. system were introduced. The burst for the two subcarriers were placed alternately on the backporch of the line synchronizing waveform in alternate lines.

Mixed-highs matrixing

Another problem in the transmission process which has to be discussed is the addition of the mixed highs to the detected colour signals in the receiver. It may be asked how the addition of high frequencies of the luminance signal must be arranged for a particular frequency limitation of the colour signal. The answer can be deduced from the requirement that for a black-and-white picture the three display signals must equal the luminance signal.

If the effect of the overall video characteristic for the red signal on this signal is denoted by the operator W , then, the signal $W E_R$ is available at the receiver. For a monochrome picture this signal is $W E_R = W E_Y$, because in that case $E_R = E_G = E_B = E_Y$. For a monochrome picture, mixed highs have to be added in the receiver to $W E_R = W E_Y$ in such a way that the red display signal " E_R " equals the luminance signal. However a certain delay with respect to the signal E_Y at the transmitter can be tolerated. Therefore the display signal for a black-and-white picture can be expressed as $V E_Y$, where operator V denotes delay. It follows from this that the high frequencies to be added are defined by

$$(V - W) E_Y \quad (84)$$

where the delay V must be specified more precisely. The effect of this delay is shown in fig. 46a. The figure shows a transient in a black-and-white picture; the corresponding colour signal, added highs and the display signal are represented. The bandwidth limitation is assumed to be of the simplest form, viz. a band limitation by an RC element. Two values of the delay are considered, zero delay and a delay of RC sec. In both cases the display signals are the same, though one lags slightly behind the other.

In a colour picture, however, in general the transients in E_R and in E_Y are of unequal height and an error will occur in the transition. See fig. 46b. In principle this error only effects the colour and does not give rise to a luminance error. The error is proportional to $(V - W)E_Y$. For a coloured picture, therefore the picture fidelity in transients depends on the delay V . If this delay is zero ($V = 1$) the whole distortion is on one side of the transient. For a delay of RC sec errors occur on both sides of the transients of opposite sign and of about the same magnitude. The latter errors will be less disturbing.

It follows from the foregoing that in $(V - W)$ an optimum value of V can

be chosen, such that the visible effect of the bandwidth reduction the colour signals have undergone is at a minimum.

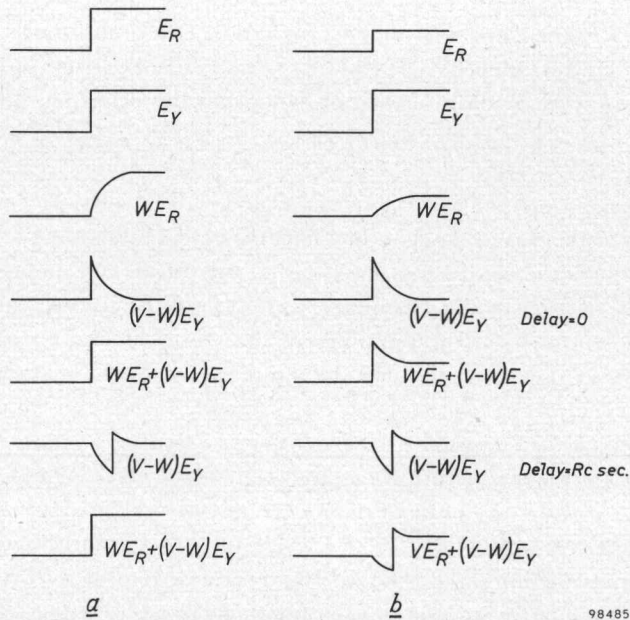


Fig. 46. The addition of the mixed highs derived from the luminance signal (E_Y) to the colour signal (E_R). The bandwidth limitation of the colour signal (W) is assumed to correspond to a single RC filter. Two values of the luminance signal delay are represented, delay = 0 and delay = RC sec. In fig. 46a the picture is monochrome ($E_Y = E_R$); fig. 46b refers to a colour picture.

In practice the addition of mixed highs can be achieved by two different methods. In the first place we may try to derive $(V - W)E_Y$ with the aid of one filter. In the second place the signals $-WE_Y$ and VE_Y may be formed separately¹⁵⁾. The bandwidth reduction W to which the colour signal in the entire transmission process has been subjected, must be exactly known in both cases. It is evident that this will be rather difficult where many stages in transmitter and receiver influence the colour-signal characteristics. For this reason it is good practice to design the transmission chain in such a way that the ultimate transmission characteristics for the colour signal is determined by simple well-defined circuitry in the receiver, preferably by a single low-pass filter after detection. If the bandwidth of the transmitter and early stages of the receiver are all wide enough, the bandwidth of the colour signals will be known exactly. Matching of the high frequencies is then much easier, especially when the second method of deriving $(V - W)E_Y$ is used.

In regard to the problem of adding mixed highs to the colour signals, we

should like to point out that the transmission of colour-difference signals represents an advantage of the N.T.S.C. system. Detection of the subcarrier provides the signal $W(E_Y - E_R)$. Hence the addition of VE_Y produces the proper red display signal, in which mixed highs and frequency limitation of E_R are matched automatically.

Comparison

Apart from mixed-highs matrixing and points mentioned above, the two-subcarrier system has been compared with the N.T.S.C. system in respect of various other aspects of the transmission problem, such as signal-to-noise conditions for the signals transmitted and the influence on picture quality, compatibility, the effect of various forms of distortion in the transmission, and the simplicity and reliability of receiver circuitry. Not only experimental equipment was built for this purpose, but also complete receivers in which many of the practical requirements for domestic colour-receivers were taken into account. These receivers have been tested at a range of some kilometers of a low-power transmitter broadcasting an experimental programme. Although the problems were analyzed, investigated and solved only in first approximation the results were acceptable.

In general it can be said that the experiments with the two-subcarrier system were successful but did not lead to the conclusion that the two-subcarrier system is superior to the N.T.S.C. system. Certainly the simple circuitry can be seen as an advantage; however, in respect of the statistical distribution of subcarrier amplitudes, the matrixing of mixed highs, and certain other points, the two-subcarrier system is at a disadvantage. After these results had been obtained, the two-subcarrier system submitted earlier to the C.C.I.R. as a possible method for colour-television transmission was — also for reasons of standardization — withdrawn as such.

III.4.3. *Other transmission systems*

Apart from the N.T.S.C. system and the two-subcarrier system there are two other proposed colour-television transmission systems which are based on the principles dealt with in III.3.

In the first place there is the R.C.A. colour system. This system was the first simultaneous compatible system developed. Originally it was based on a triplex dot-interlaced transmission of the primary signals (green, red and blue). As has been pointed out in section III.3.4, this means that the three signals are transmitted in the same way and that the process is most suited to signals having similar characteristics. Hence the system was not very well adapted to the typical characteristics and requirements of colour television. Gradually these characteristics and requirements came to be recognized. Suitable transmission principles as subcarrier transmission, constant luminance, "orange-cyan wideband trans-

mission" were invented, and the R.C.A. system was modified in accordance with these. Ultimately the N.T.S.C. system came to existence as a modified form of the R.C.A. system, though it differed considerably from the original R.C.A. conception. It will be clear that the R.C.A. system as such is at this moment only of historical interest and a further description here would not serve any purpose. Details of the system can be found in the literature ³¹).

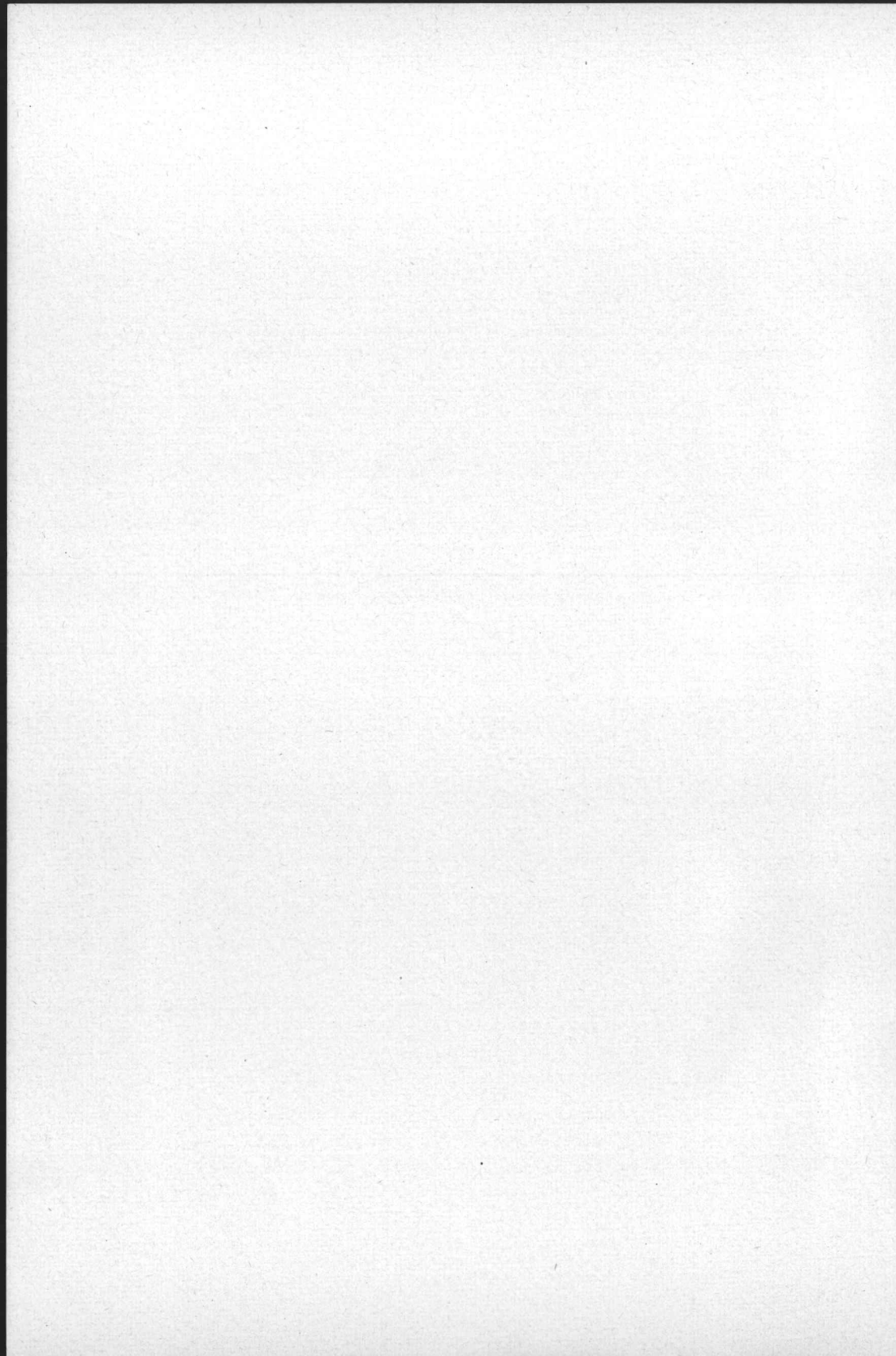
Another system in which the dot-interlace principle is employed has been developed by Boutry, Le Blan and Genève ³²); it is based on our experiments on non-synchronous detection of a duplex dot-interlace transmission. (See section III.3.3.2.) In this system the green and red primary signal are conveyed by a waveform like that shown in fig. 32d; to this wave-form non-synchronous signal separation can be applied. The blue signal with small bandwidth is transmitted on a subcarrier outside the band of the two dot-interlaced signals. This system too, of course, is open to the objections against the use of dot-interlace for colour television, as noted in III.3.4. Moreover signal-to-noise conditions are very bad, owing to the relatively large signal range occupied by the signal of fig. 32d.

Conclusion

Reviewing the development of colour television during the last ten years, we can conclude that the transmission problem has been solved in a very satisfactory way. The N.T.S.C. system has proved to be a very good and reliable system ingeniously adapted to the properties of the eye and to the typical transmission requirements. A further development of colour television as a broadcast service is hindered more by the shortcomings of existing display techniques than by difficulties in the mere transmission of the colour-television signals. Up to now the display technique employed in the R.C.A. tricolor picture tube is the only one which is being employed in large-scale production of colour television receivers. However, it can be said to be an expensive, rather complicated technique. Though other solutions have been proposed and realized, at this moment it is not very probable that the R.C.A. picture tube will be replaced by a more simple and less expensive display in a short time.

REFERENCES

- 1) J. Rajchman, Proc. Inst. Radio. Engrs **46**, 1808-1824, 1958.
- 2) Ch. Shelton a.o., J. Soc. Mot. Pict. Telev. Engrs, **67**, 441-451, 1958.
- 3) H. Lubszynski, J. sci. Instrum. **34**, 81-88, 1957.
- 4) F. Backers and J. Wessels, to be published in Philips technical Review.
- 5) R. Barthelemy, C.R. Acad. Sci., Paris **229**, 161-163, 1949.
- 6) V. Zworykin and G. Morton, Television, Wiley, New York, 1954.
- 7) L. Heijne, Acta Electronica **2**, 124-131, 1957.
- 8) C. J. Bouwkamp, Unpublished report, Philips Research Lab., 1958.
- 9) H. de Lange, Attenuation characteristics and phase-shift characteristics of the human fovea-cortex systems in relation to flicker-fusion phenomena, Thesis, Delft, 1957.
- 10) Ch. Ginsburg a.o., J. Soc. Mot. Pict. Telev. Engrs **66**, 177-188, 1957.
- 11) J. Haantjes and K. Teer, Wireless Engr **31**, 225-233, 266-273, 1954.
- 12) W. Boothroyd, Electronics **22**, 88-92, Dec. 1949. **23**, 96-99, Jan. 1950.
- 13) W. Sonnenfelt, Trans. Inst. Radio Engrs **B.T.R. 1**, 1-8, 1955.
- 14) J. Brogan, Proc. Symp. on Information Networks, N.Y. 1954.
- 15) K. Teer, Electr. Radio Engr **34**, 280-286, 1957. **34**, 326-332, 1957.
- 16) K. Teer, Unpublished report, Philips Research Lab., 1952.
- 17) L. Le Blan a.o., Onde élect. **35**, 5-21, 1955.
- 18) T. Murakami a. o., R.C.A. Rev. **16**, 580-611, 1955.
- 19) Ch. Hirsch, Advanc. Electron. **5**, 316-362, 1953.
- 20) Second color television issue, Proc. Inst. Radio Engrs **42**, 1-343, 1954.
- 21) K. McIlwain a.o., Principles of colour television, Wiley, New York, 1956.
- 22) J. Davidse, Nachr. techn. Zeitschr. **9**, 461-466, 1958.
- 23) B. Loughlin, Proc. Inst. Radio Engrs **39**, 1264-1279, 1951.
- 24) W. Bailey, Proc. Inst. Radio Engrs **42**, 60-71, 1954.
- 25) J. Haantjes and K. Teer, Wireless Engr **33**, 3-9, 1956 and **33**, 33-46, 1956.
- 26) E.B.U. Bulletin **7**, 461-486, 1956.
- 27) T. Poorter and F. de Vrijer, Philips techn. Rev. **19**, 338-355, 1957-58.
- 28) H. Seelen, R.C.A. Rev. **16**, 122-139, 1955.
- 29) R. Clapp a.o., Proc. Inst. Radio Engrs **44**, 1108-1114, 1956.
- 30) R. Dome, Electronics **23**, 70-75, Sept. 1950.
- 31) R.C.A. Rev. **10**, 504-524, 1949.
- 32) G. Boutry a.o., Onde élect. **37**, 337-357, 1957.



SUMMARY.

In this study the possibility of bandwidth compression in television transmission is considered. After a general introduction in which an outline is given and the relation between bandwidth and picture-quality in conventional television is described, an analysis is made of the redundancy present in normal television transmission. Three different aspects of redundancy are distinguished; they are referred to as the statistical aspect, the physiological aspect and the psychological aspect.

The statistical aspect is conceived with probability distributions of brightness values of picture elements. The correlation between adjacent picture elements is considered especially. The analysis is based on a practical principle of bandwidth compression. The results are compared with those obtained by some other authors.

The physiological aspect of redundancy is conceived with the properties of the eye. The properties investigated are resolving power, persistence of vision, differential sensitivity and the perception of colour. Experiments on the resolving power were carried out in order to ascertain the extent to which definition in the screen image can be matched to the place-dependent resolving power of the retina. Persistence of vision was investigated in relation to the perception of discontinuities in motion. Differential sensitivity covers all effects related to the ability to perceive differences in brightness. A description is given of special effects in this field, as well as of proposals which have been made for transmission systems in which a decrease of the required channel capacity is achieved in virtue of these effects. With respect to the perception of colour, a survey is given of the investigations which have been carried out in recent years on behalf of the colour television, and which have led to the transmission principles now in use in colour television.

Only a few remarks are made about the psychological aspect of the redundancy, because this subject is too remote from practical television engineering, and therefore unlikely to yield useful conclusions.

The third part of the study deals with transmission systems with narrow bandwidth which have been investigated by the author. These investigations were restricted to transmission systems in which bandwidth compression is effected by a decrease in the number of frames per sec. This number can be decreased in two different ways: by decreasing the field frequency or by decreasing the information per field. The necessary and sufficient condition for practical realization of the former method is a suitable memory device. Accordingly, we have restricted ourselves to considering the memory device in general and in particu-

lar, a vidicon-type camera tube which could be used as a fairly simple device of this kind.

The latter method, a decrease in the information per field, consists in the use of dot-interlace and subcarrier techniques which have been developed for the purpose of colour television. Both principles are examined in detail and a comparison is made between typical features.

Finally, the use of these principles in colour television are considered in an account of colour-television transmission systems. For the main part these considerations deal with the N.T.S.C. system and a two-subcarrier system into which extensive investigations have been made by the author.

Acknowledgements

The investigations underlying this thesis have been carried out in the Philips Research Laboratories, Eindhoven, Netherlands. I am greatly indebted to the Directors of the laboratories for enabling me to carry out this work, in particular to Dr J. Haantjes under whose stimulating guidance I became acquainted with television techniques and television problems. In addition I want to express my gratitude to my colleagues, especially to Mr J. J. P. Valetton, Dr F. W. de Vrijer and Mr G. Lubben, for their contribution in equipment and knowledge, for their critical comments and in general for the close cooperation we had. Further, I wish to thank Messrs. L. M. Greep, N. J. P. Klijn and G. Menkveld for their assistance in the development and construction of equipment, Mr. A. F. Monypenny for his assistance in translating the original text into English, and the administration of the laboratories for their help in preparing the manuscript.



Fig. 6. Picture for which the percentage of "essential picture elements" has been measured.
See p. 19.



Fig. 7. Picture for which the percentage of "essential picture elements" has been measured.
See p. 19.

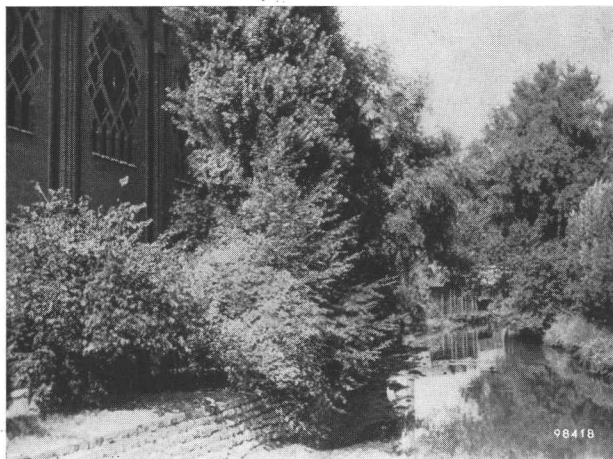


Fig. 8. Picture for which the percentage of "essential picture elements" has been measured.
See p. 19.



Fig. 12. Television picture in which the definition is matched to the place-dependent resolving power of the retina. See p. 30.



Fig. 13. Television picture in which the definition is matched to the place-dependent resolving power of the retina. See p. 30.

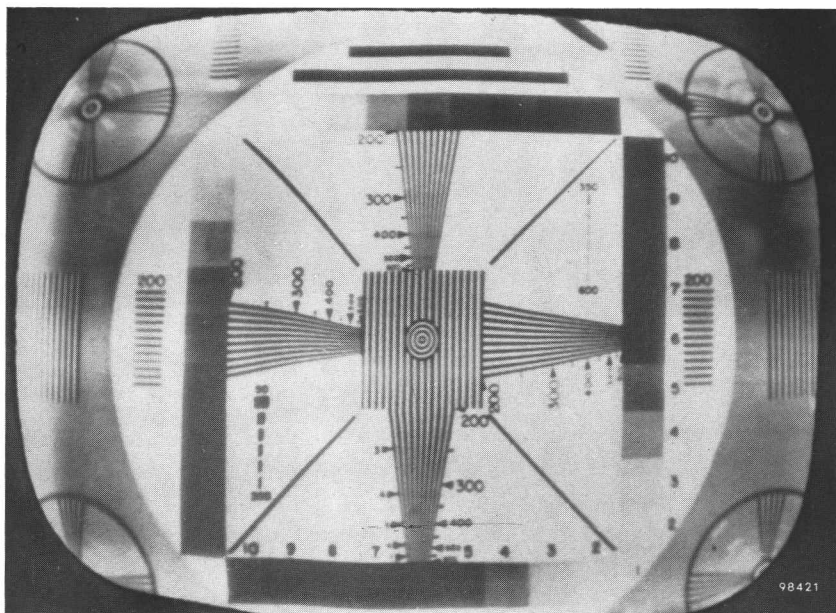


Fig. 36. Test pattern reproduction for normal channel bandwidth of 5 Mc/s. See p. 100.

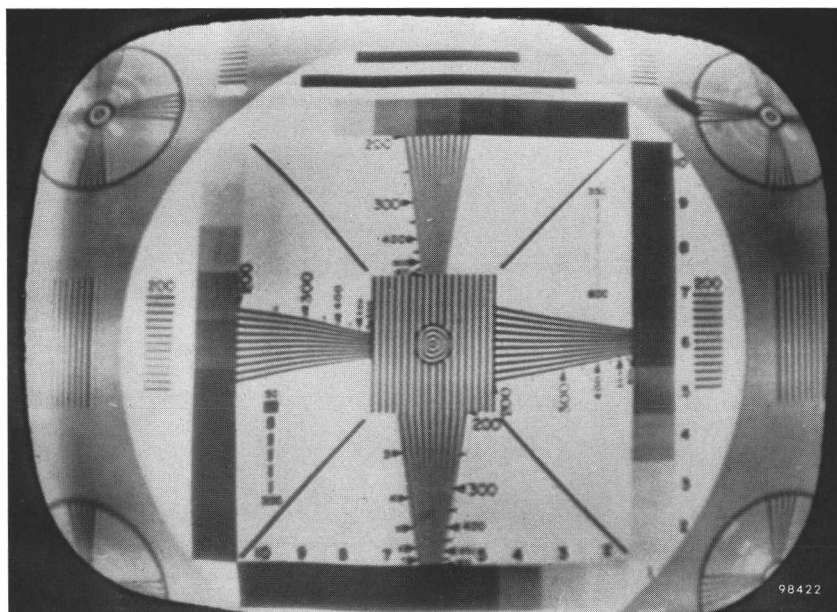


Fig. 37. Test pattern reproduction for half the normal channel bandwidth of 2.5 Mc/s. See p. 100.

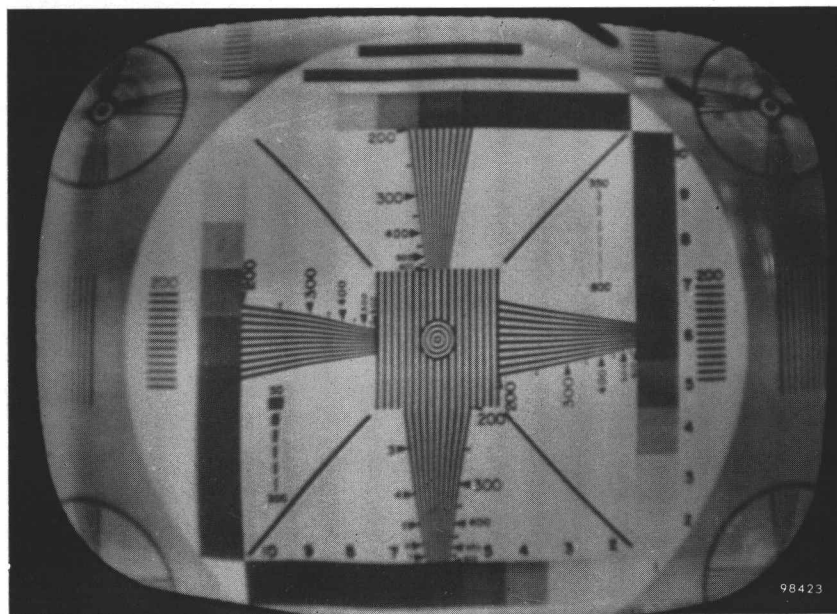


Fig. 38. Test pattern reproduction for dot-interlace transmission with a channel bandwidth of 2.5 Mc/s. See p. 100.

STELLINGEN

I

Het moet zeer onwaarschijnlijk geacht worden, dat de correlatie tussen de helderheidswaarden van naburige beeldelementen, dan wel de correlatie tussen helderheidswaarden van een beeldelement in opeenvolgende aftastingen, tot een televisieoverdracht met verminderde bandbreedte aanleiding kan geven, welke te prefereren is boven een overdracht waarbij eenzelfde verminderde bandbreedte is bereikt door een frequentiebegrenzing van het normale televisiesignaal zonder meer.

Dit proefschrift, II.2.1.

II

Hoewel de keuze van vierentwintig beeldjes per seconde in de normale filmtechniek zeker een verantwoorde keuze is, kan toch worden gesteld, dat een vergroting van dit aantal nog significant bijdraagt tot verhoging van de werkelijkheidsindruk van het gereproduceerde beeld.

Dit proefschrift, II.3.2.1.

III

De beschouwingen van Krasil'nikov ten aanzien van de waargenomen signaal-ruis verhouding in televisiebeelden kunnen tot verkeerde conclusies aanleiding geven omtrent het verschil in hinderlijkheid van stationnaire en van veranderende storende patronen.

Krasil'nikov, Radio Engineering, **12**, 15-23, 1957.

IV

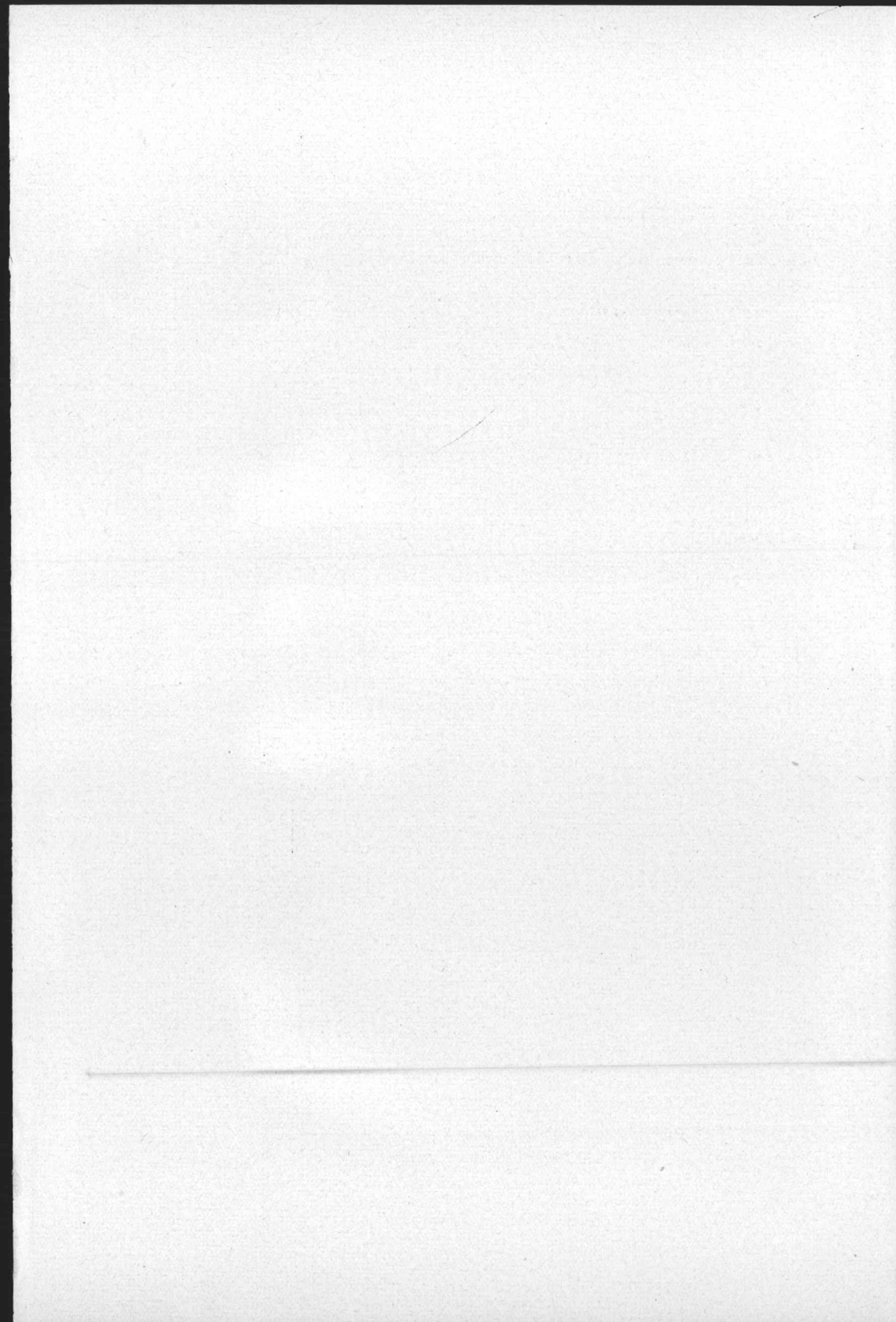
Voor vierpolen opgebouwd uit een eindig aantal passieve netwerkelementen is, afgezien van duidelijk triviale gevallen, geen configuratie te vinden waarbij de overdracht van een serie oneindig korte impulsen met volkomen onderlinge onafhankelijkheid mogelijk is.

Dit proefschrift, III.3.1.1.

V

De veelal gebezigde analyse van de televisieoverdracht aan de hand van het frequentiespectrum van het televisiesignaal geeft slechts gebrekkige inzichten omtrent de mogelijkheden om tot een efficiënter gebruik van het televisiekanaal te komen. Dit geldt in het bijzonder ten aanzien van de toepassing van het interpuncterings- en het hulpdraaggolfprincipe.

P. Merz en F. Gray, Bell Syst. tech. J., **13**, 464-480, 1934.
J. Abrahams, Proc. Inst. Radio Engrs, **42**, 81-83, 1954.
H. Schönfelder, Frequenz, **10**, 1-5, 1956.



VI

De voorwaarde, dat een kleurentelevisiesysteem verenigbaar dient te zijn met de bestaande methode van zwart-wit televisie, sluit slechts weinig mogelijkheden voor de transmissietechniek uit, die gelet op de kleurentelevisieoverdracht als zodanig als gunstiger oplossing kunnen worden aangemerkt.

Dit proefschrift, III.4.

VII

Het realiseren van een z.g. „modulatie met onderdrukte draaggolf” zonder balansschakeling is in principe mogelijk door gebruik te maken van een hogere harmonische van een geschikt gekozen periodiek signaal waarop een gemoduleerde begrenzing is toegepast.

VIII

Bij de magnetische registratie van stapfuncties kan met voordeel gebruik gemaakt worden van overeenkomstige correctiemiddelen als in de televisietechniek voor het vergroten van het scheidend vermogen van opneembuizen wordt toegepast. Een variatie ten opzichte van de uitvoeringsvormen voor laatstgenoemde toepassing is noodzakelijk om de registratie van transversale componenten mede in rekening te brengen.

IX

In diverse gebieden van de telecommunicatietechniek zou een eenvoudig signaalvertragend systeemelement een nuttige toepassing kunnen vinden. De mogelijkheden om een tijdsvertraging in bedoelde vorm te realiseren zijn nog onvoldoende onderzocht.

X

In de psychologische achtergrond van het toegepast natuurwetenschappelijk onderzoek kunnen zeker ludieke elementen worden onderkend. De beschouwingen van Huizinga op dit punt zijn onvolledig.

J. Huizinga, *Homo Ludens*, Tjeenk Willink, Haarlem, 1938.

XI

Voor een serie aequidistante puntvormige stralers of ontvangers geplaatst op één lijn kan, indien de onderlinge afstand voldoende klein is ten opzichte van de golflengte, een doeltreffende in de microfoontechniek bruikbare methode worden aangegeven, om de signalen der elementen naar grootte en frequentieafhankelijkheid te bepalen voor de synthese van frequentieonafhankelijke richtingsdiagrammen, in het bijzonder van het diagram met maximale richtingsgevoeligheid. Wat dit betreft zijn de beschouwingen van Pritchard onvolledig.

R. Pritchard, *J. acoust. Soc. Amer.*, 26, 1034-1039, 1954.

XII

De wijze waarop de financiering van het lager onderwijs wettelijk is geregeld leidt tot ongewenste discrepanties in onderwijsmogelijkheden zowel tussen scholen in verschillende gemeenten als tussen scholen van verschillende grootte.

Lager-Onderwijs Wet 1920, Art. 54, 55 en 101.

XIII

Het verdient aanbeveling om bij de ingenieursopleiding meer aandacht te besteden aan de vaardigheid waarmee de betrokkenen zich ten aanzien van hun vak in woord en geschrift weten uit te drukken.

XIV

Naast een zuiver amuserende doelstelling kan de zogenaamde kleinkunst een waardevolle functie toegekend worden als uitingsvorm voor commentaar en critiek op het maatschappelijk en politiek gebeuren. Het valt te betreuren dat deze mogelijkheid, waarmee een activerend element in het vermaak wordt gebracht, onvoldoende wordt uitgebuit in het bijzonder bij de radio- en televisie-omroep.