RESEARCH DEPARTMENT

SOME EXPERIMENTS ON THE SUBJECTIVE EFFECT OF LIMITING THE UPPER FREQUENCY RANGE OF PROGRAMMES

Report No. L-034

(1957/ 10)

D.E.L. Shorter, B.Sc.(Eng.), A.M.I.E.E.
W.I. Manson, B.Sc.(Eng.),
E.R. Wigan, B.Sc.(Eng.), A.M.I.E.E.

(W. Proctor Wilson)

SOME EXPERIMENTS ON THE SUBJECTIVE EFFECT OF LIMITING THE UPPER FREQUENCY RANGE OF PROGRAMMES

May 1957

Report No. L-034

(1957/10)

PRIVATE AND CONFIDENTIAL

SOME EXPERIMENTS ON THE SUBJECTIVE EFFECT OF LIMITING THE UPPER FREQUENCY RANGE OF PROGRAMMES

SUMMARY

To obtain first-hand knowledge of the subjective effects of high-frequency band limitation, some tests have been carried out with Research Department staff of various categories as observers (thirty-four in all), using twelve types of recorded programme material.

Low-pass filters having nominal cut-off frequencies of 10 kc/s and 12 kc/s were used to restrict the band, as well as a circuit giving a slowly increasing attenuation at frequencies above 7 kc/s.

The observer was presented with a ten-second passage of programme played three times in succession, in each case with one of two different band-widths. The scoring was based on the observer's ability to tell which two of the conditions presented were alike. Four such tests were carried out on each observer for each programme item and the whole process was repeated for each circuit condition to be studied.

The data obtained with experienced and inexperienced observers are analysed separately. The significance of the results is discussed and suggestions made for further experiments.

The series of experiments described, which has given results of great interest, has been irreverently but not inaptly denominated "Operation Cloth-Ear".

1. INTRODUCTION

The upper limit of the frequency range required for the accurate reproduction of the various musical and other sounds which may occur in broadcast or recorded programmes has long been known to lie in the region of 15 kc/s. Until a decade or so ago, the defects of commercial disk recordings and medium-wave broadcast reception limited the band available to the general public to one half or even one quarter of this figure. With the advent of v.h.f. broadcasting and of various improvements in recording, the reproduction of the full high-frequency range under domestic conditions has at last become a technical possibility. In broadcasting, however, an economic limit to the frequency band radiated is set by the high cost of land-line transmission, and some compromise between fidelity and expense is therefore necessary.

To make any rational decision on the frequency range to be transmitted, it is necessary to have some idea of the subjective effects produced by various degrees of band limitation, and several investigations designed to produce this information have been described in the literature. However, the results obtained from experiments of this kind depend so much on such factors as the age and experience of the observers, the nature of the programme material and the method of presentation that caution is necessary in applying the data given. In December 1955, a demand arose for an investigation within the Corporation into the effects of band restriction observed under well-defined conditions; in particular, it was urgently desired to discover what degree of degradation in programme quality would result if the upper frequency range were to be limited to approximately 10 kc/s. Experiments were accordingly carried out using low-pass filters designed to produce attenuation increasing either slowly or rapidly with frequency in this region. After examination of the results obtained, it was decided to undertake further tests, using a 12 kc/s low-pass filter. It was recognised that the effect of this filter would be in some measure obscured by the combined losses, at frequencies above 12 kc/s, of the recording system, the loudspeaker and, in some cases, the microphone. Since, however, any improvement in instrumentation, even if possible at the time, would have prevented direct comparison with the earlier experimental results, it was decided, with the reservations made above, to employ the same experimental equipment and test material as before.

Because of the large number of observations required in subjective tests to give a statistically significant result, and the necessity to obtain results quickly, the number of variables in this experiment was restricted by an arbitrary choice of observers and of programme material. The probable effect of these restrictions and the considerations involved in extending the scope of the investigation will be discussed later.

2. CONDITIONS OF EXPERIMENT

2.1. Filters used to limit High-Frequency Range

Fig. 1 shows the frequency characteristics of the low-pass filters used to limit the high-frequency range. Filters I and II were chosen to illustrate the difference between the effects of slow and rapid cut-off. Filter I, which has an attenuation of 2½ dB at 7 kc/s and 12 dB at 10 kc/s, is a single constant-k section giving an ultimate rate of change of attenuation of 18 dB per octave. Filter II is a multisection commercial product with an attenuation, relative to the mid-band value, of 2½ dB at 10 kc/s and more than 20 dB at 11 kc/s. In the last set of experiments, a further multisection Filter III, attenuating 2½ dB at 12 kc/s and more than 25 dB at 13 kc/s, was employed.

2.2. Observers

In considering the choice of observers it is necessary to enquire at the outset whether the opinion of the whole listening public is to be sought or that of the most critical section of the population. Since the answer to this question must depend on the purpose for which the information is required and on considerations of policy with which this report is not concerned, it is desirable to obtain data from observers representing both categories. It is not difficult within the Corporation to find a sufficient number of critical listeners to give statistically significant results; to obtain an equally representative sample of the general public is, however, much less simple. Because of the urgency of the work, it was necessary to restrict

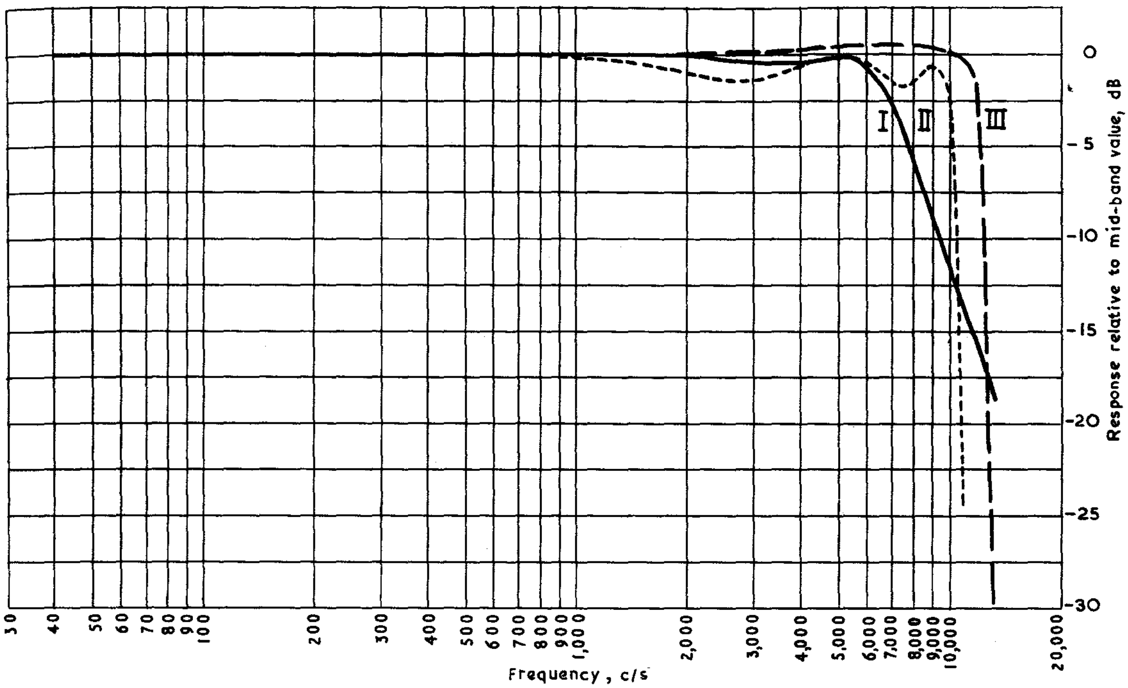


Fig. 1 - Frequency characteristics of Filters I, II and III

the scope of the investigation to a number of Research Department staff whose services could be obtained conveniently at the time. Of these, a group of twelve engineering staff experienced in critical listening was taken to represent, for the time being, the most discriminating section of the listening public. The remainder, numbering twenty-two, who will be referred to for convenience as "inexperienced observers", were mostly non-technical or semi-technical; this group was drawn from as many different categories of staff as possible, but could not of course be described as a representative sample of the population as a whole.

It was decided in the first instance to confine the experiment to observers whose hearing was not seriously impaired by age and to this end an arbitrary upper age limit of forty years* was imposed. No otological examination of observers was carried out but individuals known to be hard of hearing were excluded.

The various categories of staff were represented as follows:

EXPERIENCED AND TECHNICAL OBSERVERS

12 Engineering staff from Electro-Acoustics Group (male)

INEXPERIENCED OR NON-TECHNICAL OBSERVERS

6 Engineering staff from Electro-Acoustics Group without extensive listening experience.
1 Graduate Apprentice.
5 Secretaries.
2 Workshop mechanics.
1 Laboratory mechanic.
7 Drawing Office staff.

(12 male, 10 female)

*It may be noted that the standards of "normal" hearing used in otology are based on measurements made with observers between eighteen and twenty-five years of age.

2.3. Loudspeakers

The loudspeaker employed was of an experimental type in which the axial response was maintained within 4 dB of the mid-band value to 12 kc/s and within 9 dB up to 15 kc/s. As a compromise to the ideal of having all observers "on axis" three observers at a time were arranged, at a distance of 1·5 m from the loudspeaker, in an arc subtending an angle of some ± 15° about the loudspeaker axis; the maximum additional high-frequency loss arising from the directional properties of the loudspeaker was then only 5 dB at 12 kc/s.

2.4. Programme Material

The selection of programme material for a test of this kind could logically be carried out in one of two ways. On the one hand, a representative cross-section of all types of programme might be sampled, thus weighting the data in favour of the most frequently occurring sounds; the results would in effect be based on the average spectrum of the programme. Apart from the fact that the selection of material for such a survey would involve a number of highly arbitrary decisions, arguments based on averages are particularly dangerous when applied to aesthetic matters. As a *reductio ad absurdum*, it might be contended, on such a premise, that the triangle part could safely be omitted from an orchestral composition on the grounds that the instrument is played for only a small proportion of the programme time. The alternative and more appropriate procedure followed in these experiments is to consider only the most critical programme material. In this case a single test passage would in theory suffice; in fact, the various items used were found by a pilot test to have the same order of susceptibility to band restriction, so that the experiment was in principle the same as if only one type of material had been used throughout.

The test programme was recorded on magnetic tape and consisted of twelve items which had been selected, on the basis of pilot experiments, as being most likely to show up the effects of band limitation. The following sounds were included:

Tubular bells.
Orchestral brass combination.
Orchestral cymbal.
Dance band with cymbals.
Snare drum.
Military drum.
Samba band including maraccas, handclaps and triangle.
Female speech.

Some of the above items were selected from broadcast programmes; others were dubbed from some instructional recordings of individual instruments which were being made by Operations and Maintenance Department at the time. In most of the items, the microphone employed was either the ribbon type PGS/1 or the electrostatic type C.12, both of which have an axial frequency response extending to 15 kc/s[1,2]. For most of the dance band items, in which a number of microphones of the older ribbon types AXB or AXBT are often used, arrangements were made with the studio manager concerned to substitute a PGS/1 or C.12 microphone in front of the solo instrument. In a few instances, material picked up by AXB microphones alone had to be used; in such cases, the high-frequency content of the reproduced programme, while sufficient to show the effect of band limitation, was lower than would be obtained with modern microphones[1].

The frequency response of the tape recorders showed no loss at 12 kc/s but up to 2½ dB loss at 15 kc/s. Since, in preparing the test tape, the original material had to be edited and re-recorded, in the worst case this loss could have been doubled.

2.5. Noise Level

To give a maximum flexibility in the experimental arrangements, the programme material was recorded without intentional restriction of the band-width, the necessary filters being introduced as required on replaying. It was found in preliminary trials that the presence of a low-pass filter could often be detected by its effect on the tape noise, even when the effect on the recorded material was imperceptible. To avoid the use of this artefact by the observers as an aid to discrimination, a small and constant quantity of white noise, sufficient to mask the tape noise, was injected into the circuit after the point at which filters were inserted.

2.6. Sound Level

The sound level presented to the observers on peaks of programme was in the region of 70 to 85 phons.

2.7. Presentation

In the first experiment, the effect of Filter I was compared with that of the unrestricted band. A second, similar, experiment was carried out with Filter II, after which a direct comparison was made between Filters I and II. Finally, the effect of Filter III was compared with that of an unrestricted band.

The method of presenting the test material and of ascertaining opinions was devised so that no technical understanding on the part of the observer was called for, while the effect of any predisposition of an individual in favour of either a wider or a narrower frequency band was as far as possible eliminated.

For each test, a sample of recorded programme approximately ten seconds in duration was presented three times with intervals of two seconds; these three presentations were referred to as A, B and X. A and B represented the two conditions to be compared; for example, one of them might be presented with the full frequency range of the equipment while the other had a restricted frequency band. Condition X was identical with either condition A or condition B. The observers were told before the commencement of the tests that A and B would differ and that X would be identical with one of them. They were then asked to record their opinions in the form "X is like A", "X is like B" or "don't know", the last-named possibility being allowed on psychological grounds to avoid over-anxiety. It would have been impossible to conceal from all the observers the nature of the difference between A and B; it was therefore decided, for the sake of uniformity, to make this known to all by a preliminary demonstration, with a 6·8 kc/s low-pass filter, the effects of which were so pronounced as to make any comment redundant.

It should be noted that the test described is very much less critical than the more usual procedure of allowing a quick change of condition during the programme, for unless the difference between B and X is obvious, the observer has to make use of his memory. In some cases, the observers reported that A and B were obviously different, but that it was still not obvious which of the two most resembled X.

Each pair of band-widths to be compared was the subject of a separate experiment. In each experiment, the recorded material was arranged to give an A, B, X presentation twice in succession on each item of programme, an interval of some five seconds being allowed between the two tests. This procedure, applied to the twelve selected programme items, required about twenty minutes, but to avoid excessive fatigue on the part of the observers, the test material was divided into two parts, each of about ten minutes' duration, which were presented in succession with an interval of five minutes. The whole process was repeated after an interval of not less than half an hour, in some cases on the following day. Thus, in the comparison of each pair of band-widths, every observer had forty minutes' total listening time, every one of the twelve programme items was the subject of four A, B, X tests and the total number of judgments per observer was forty-eight.

With every passage of programme each of the four possible orders of presentation in the A, B, X test (e.g., A and X band-restricted, B full range; B and X band-restricted, A full range ... and so on,) was employed on an equal number of occasions, the different arrangements being used in random order.

3. RESULTS

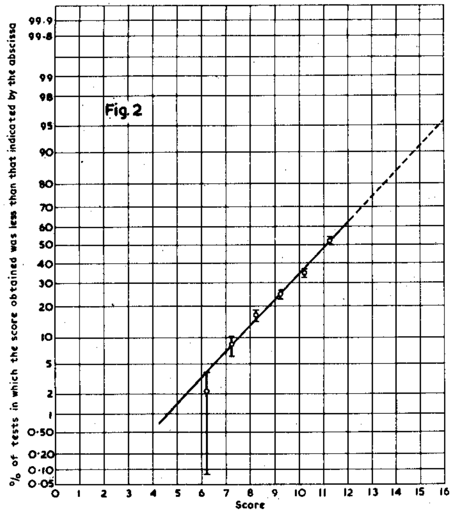
3.1. Method of Analysis

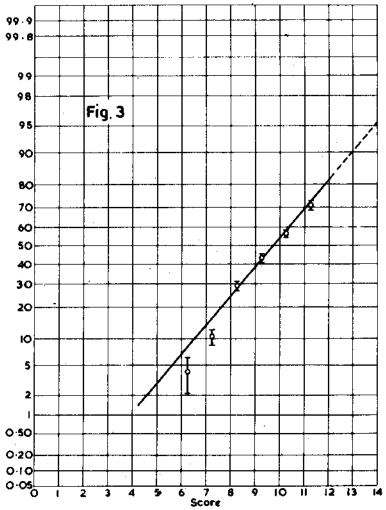
In the discussion which follows, the term "score" refers to the number of "right" judgments recorded, i.e. to the number of occasions on which the observer correctly described the test condition presented to him. In allowing the observer to return a non-committal answer if he wished, a number of judgments was obtained under the heading "don't know", which would otherwise have been distributed at random between "right" and "wrong". In analysing the results, the answers appearing under this heading have been utilised by dividing them equally between the two categories; as a result of this process, some of the final scores contain half units.

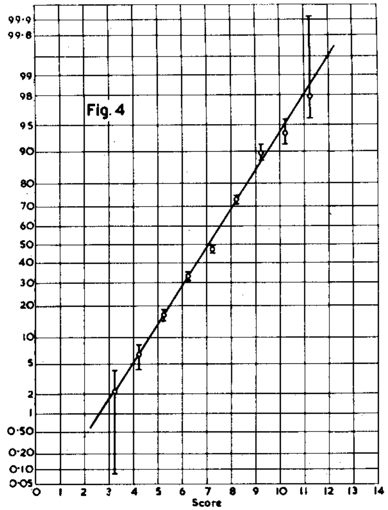
3.2. Presentation of Data

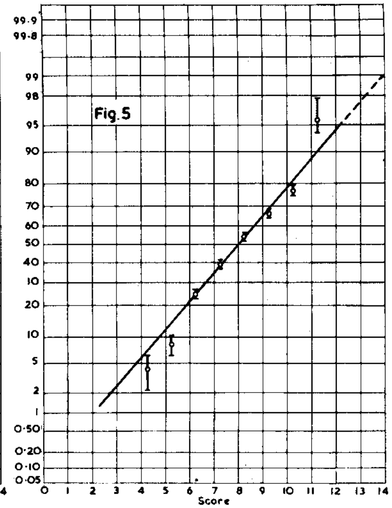
The scores recorded for each observer in the various tests are plotted in Figs. 2 to 7 in the form of cumulative frequency ordinates on an arithmetic probability scale. Each ordinate shows the percentage of tests in which the score was less than the number indicated by the corresponding abscissa.

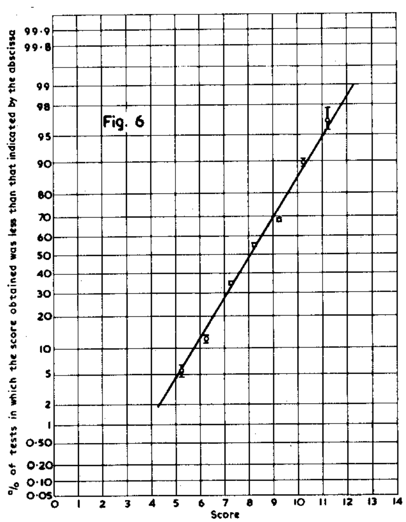
The number of different programme items presented was twelve; this number therefore represents the maximum score obtainable by a single observer for all items taken together. Since the entire experiment was performed four times for every test condition, the total number of such scores registered in each case was four times the number of observers. The possible score values all fall on whole or half units; to avoid ambiguity, therefore, the class divisions are carried out at the quarters, so that the ordinates represent the number of scores "less than $1\frac{1}{4}$", "less than $2\frac{1}{4}$" and











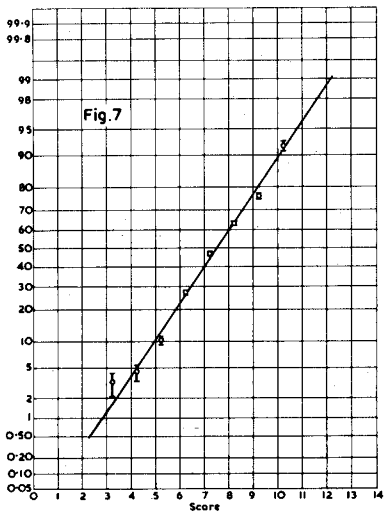


Fig. 2 - Experienced Observers. Comparison: Filter I/Unrestricted Band

Fig. 3 - Experienced Observers. Comparison: Filter II/Unrestricted Band

Fig. 4 - Experienced Observers. Comparison: Filter III/Unrestricted Band

Fig. 5 - Experienced Observers. Comparison: Filter I/Filter II

Fig. 6 - Inexperienced Observers. Comparison: Filter I/Unrestricted Band

Fig. 7 - Inexperienced Observers. Comparison: Filter II/Unrestricted Band

so on up to "less than 12½".

It is a property of the probability scale employed in Figs. 2 to 7 that the distance representing a given percentage change is a minimum in the region of the 50% mark but increases as the reading approaches 0% or 100%; in drawing the line of best fit, therefore, a greater divergence may be permitted to points lying near the two extremes of the scale. To allow due weight to be given to every point in drawing the line, each one has been marked with upper and lower limits showing the effect of a fixed change — in the present case, plus or minus one score unit. An analogous procedure was described in an earlier Research Report[3].

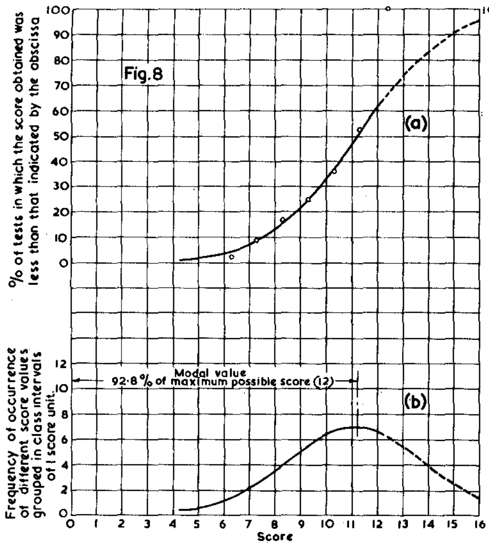
Figs. 8(a) to 13(a) show the data of Figs. 2 to 7 replotted on a linear ordinate scale, so that each straight line becomes a curve of sigmoid form.

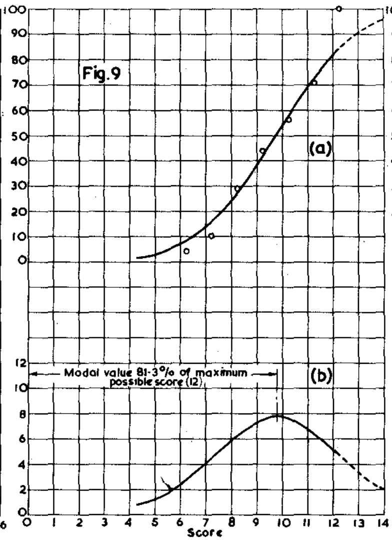
The straight lines in Figs. 2 to 7, and hence the corresponding sigmoid curves of Figs. 8(a) to 13(a), relate to the normal Gaussian frequency distribution, according to which any value of score, from zero to infinity, is possible (though all values are not, of course, equally probable). In practice, there must always be some finite maximum value of score — in the present case twelve. The practical limitation thus imposed can usually be ignored as long as the proportion of observers who achieve the maximum score is small. If, on the other hand, more than about 25% of the observers obtain the maximum possible score, the situation illustrated in Figs. 8(a) and 9(a) results. It will be seen that an appreciable part (shown dotted) of the sigmoid curve lies above 12½ on the abscissa scale; at the same time, the point representing the number of scores "less than 12½" stands well above the curve. The reason for these apparent anomalies is that some of the observers who scored twelve were potentially capable of performing a more difficult task than that set and would therefore have scored more than twelve had this been possible. The type of distribution shown in Figs. 8(a) and 9(a) could equally well have been obtained from a test in which the maximum score was higher than 12, if, by an arbitrary decision, all scores over 12 were subsequently cut down to that figure. The effect of "compressing" the data in this manner is to increase the number of scores in the highest category recorded — in the present instance, "less than 12½", while leaving the other experimental points unchanged.

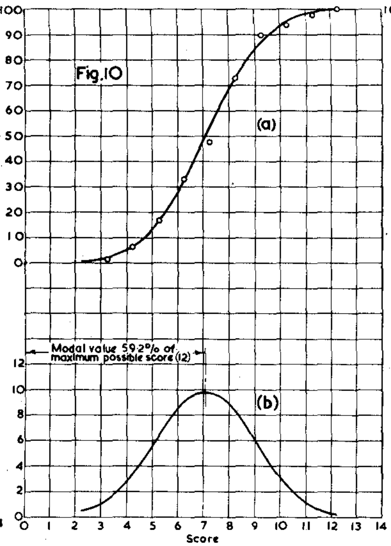
To illustrate the practical significance of the scores obtained, let it be supposed that all the ten-second test passages used in any one experiment were joined together to form a continuous programme. Then it may be said that an observer who scored X% of the possible maximum was able to give "correct" judgments for X% of the programme time.

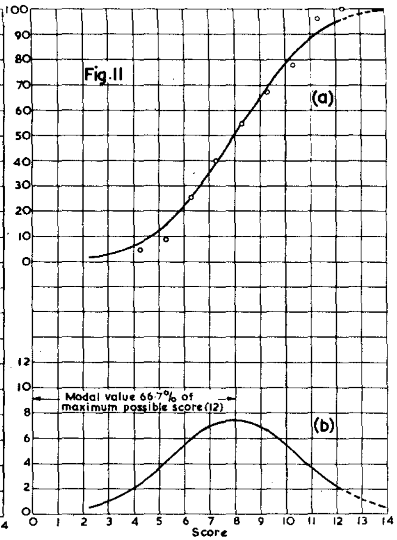
Figs. 8(b) to 13(b) show the histograms corresponding to the cumulative frequency curves of Figs. 8(a) to 13(a). The quantity of principal interest in these histograms is the modal value — i.e., the most commonly occurring value — of the score, as indicated by the Gaussian curve. This quantity represents, in effect, the answer most frequently given by observers in reply to the question "For what proportion of the programme time were you able to identify correctly passages having the same upper frequency limit?".

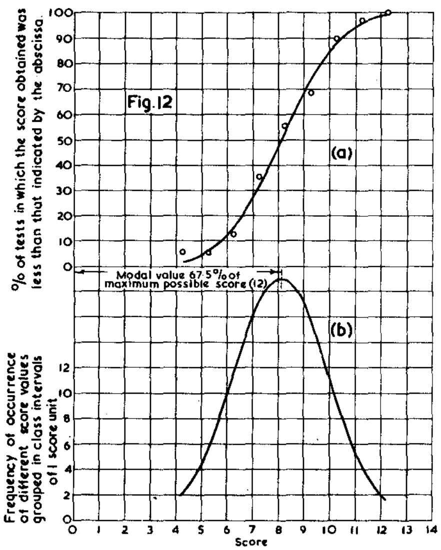
The modal values of the scores obtained from the histograms are summarised in the following table. In interpreting the results, it must be remembered that 50% of "correct" judgments could be obtained by chance alone, i.e. that a 50% score would indicate that the observer was unable to distinguish between the different conditions presented to him.











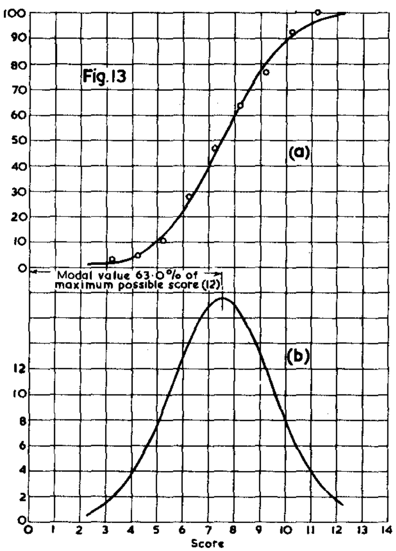


Fig. 8 - Experienced Observers. Comparison: Filter I/Unrestricted Band

Fig. 9 - Experienced Observers. Comparison: Filter II/Unrestricted Band

Fig. 10 - Experienced Observers. Comparison: Filter III/Unrestricted Band

Fig. 11 - Experienced Observers. Comparison: Filter I/Filter II

Fig. 12 - Inexperienced Observers. Comparison: Filter I/Unrestricted Band

Fig. 13 - Inexperienced Observers. Comparison: Filter II/Unrestricted Band

CONDITIONS COMPARED	MODAL VALUE OF SCORE EXPRESSED AS % OF POSSIBLE MAXIMUM	
	12 Experienced Observers	22 Inexperienced Observers
Filter I/unrestricted band	92·8 ± 6·8	67·5 ± 3·2
Filter II/unrestricted band	81·3 ± 5·8	63·0 ± 3·6
Filter III/unrestricted band	59·2 ± 4·6	-
Filter I/Filter II	66·7 ± 6·1	-

In experiments of this kind, in which the number of observers is small, it is essential to make some estimate of the stability of the data which they yield, for it is highly probable that a somewhat different result would be obtained if each test were to be repeated with a different set of observers having the same degree of experience or inexperience, or even with the same observer again. The resulting degree of uncertainty is expressed by giving, instead of a single figure for the modal value of the score, a range of values deduced by known statistical methods from the standard deviation of the results and from the number of observations made. The range of values shown in the present results are those lying within the 95% confidence limits, i.e., there is estimated to be a 95% probability that the mean* of all results obtained by repeating the experiment a very large number of times would fall within the limits given. The narrower the range of values covered by the confidence limits, the higher the degree of statistical significance of the results.

Judging from the difference between the results obtained with the experienced and inexperienced observers in the first two experiments shown in the table, it was not thought worth while to repeat the last two experiments with observers in the latter category, as it seems unlikely that the score would be significantly above 50%.

From the above results it will be seen that the presence of Filter II is apparent to a number of inexperienced observers and that of Filter III to a number of experienced observers; the presence of Filter I is naturally more noticeable, to members of both groups, than that of Filter II. The modal value of the scores for the Filter I/Filter II comparison lies well below that obtained in the comparison Filter I/unrestricted band, whence it appears that an appreciable part of the loss produced by Filter I is associated with components of the signal having frequencies above about 10·5 kc/s. On the other hand, the scores obtained with Filters I and II taken individually do not differ as much as might be expected. The results taken as a whole tend to suggest that *the frequency band from 8 kc/s to 10 kc/s may be less important to the listener than the 10 kc/s to 12 kc/s band*, but further work would be required to give certainty on this point.

In considering the foregoing results it must be remembered that the unfiltered programme material presented to the listeners was by no means unrestricted in band-width, as the frequency characteristics of the loudspeakers, the recording system, and in some cases the microphone, combined to produce an appreciable loss at frequencies above 12 kc/s. Although the experimental arrangements were probably adequate for the preliminary study of Filter II (10 kc/s) which was the primary object of the tests, the effect of Filter III (12 kc/s) is probably underrated as a result of the losses referred to.

*For the purpose of calculating the confidence limits, it is assumed that the modal and mean values of the histograms are the same, though the available data is in fact insufficient to establish identity.

4. APPLICATION OF RESULTS

The results given in the last section, while capable of serving as a practical guide, cannot provide a simple answer to the question as to whether a given degree of band limitation is permissible. The utmost that can be achieved by any investigation of this kind is an estimate of that score which would have been obtained had it been possible to test all the individuals in the category concerned, instead of having to rely on a sample. When once the score for the whole of the chosen population is known, or has been estimated with sufficient precision from a small-scale test, the concept of statistical significance, which had to be introduced because of the limited size of the sample, no longer arises. Thus, the table in the last section, while giving the best available estimates for the two classes of observers employed, should not be used as a basis for such statements as "the effect of introducing Filter II into the circuit is (or is not) significant".

The decision whether an event perceptible to a given proportion of the population is to be regarded as important in the engineering sense is necessarily an arbitrary judgment governed by considerations which cannot always be made the subject of a scientific experiment. With this reservation, the following conclusions may be reasonably drawn:

(a) The effect of restricting the high-frequency range to 12 kc/s is perceptible to only a small number of critical observers.

(b) Restriction of the frequency range to 10 kc/s is perceptible to many inexperienced observers and readily detectable by the experienced.

(c) Progressive attenuation of the response above 7 kc/s, as produced by Filter I, gives results which are evident to many inexperienced observers and obvious to most experienced observers.

5. FUTURE EXPERIMENTS

5.1. General

The experiments described in the earlier part of this report were necessarily limited in their scope. Ideally, the investigation should include other classes of observer together with a greater variety of programme material, while a larger number of observations should be taken in each set of conditions. Since, however, the data presented in this report, incomplete as it is, represents the result of over 4500 tests, it is necessary before embarking on further operations to consider carefully what extension in the scope of the investigation would most repay the effort involved. The various aspects of the work will therefore be considered in turn from this point of view.

5.2. Selection of Observers

It may well be that the observers hitherto described as "inexperienced" were in fact more experienced on the average than members of the general public. To give a representative cross-section of the population, however, many different classes of individual would have to be included and if it were desired to obtain an unambiguous result applicable to the whole body of broadcast listeners, an operation of gargantuan proportions might well be necessary*. Short of such extreme measures, it seems unprofitable to go further afield in the choice of observers. The two classes of individual employed in the present work are sufficient on the one hand to provide the most critical assessment possible for a given experimental condition and, on the other, to indicate the rate at which perception of the phenomenon being studied varies with the degree of sophistication of the individual.

5.3. Selection of Programme Material

In any future investigation it might be advisable to make a more extensive search for programme material to see whether any specially critical type of sound has been overlooked. For example, the particular recordings of orchestral strings which were made at the start of the investigation were found to be little affected by Filter II and for this reason stringed instruments were not included in the present tests, but it would be unwise without further investigation to conclude that this class of material can be safely neglected**. Moreover, it has so far been assumed that sounds of the "effects" type, other than those forming part of the percussion section of dance bands, are outside the scope of the investigation; if all such sounds were to be included, the effects of band limitation would probably be more pronounced than those so far indicated.

5.4. Selection of Frequency Bands

The three low-pass filters used in these experiments were chosen to illustrate the effect of band restriction on the reproduction of the most critical types of sound and thus to indicate the requirements for a permanent line system carrying every kind of programme. For some programmes, however, temporary lines are employed and it would be an advantage to know in each case whether the nature of the material would permit the use of a link having a restricted frequency range. To this end it might be useful to apply the methods adopted in the present experiment to recorded excerpts from sporting commentaries, public functions or other programme material for which temporary links may in future be required.

5.5. Instrumentation

Reference has been already made to some of the shortcomings of the instrumental arrangements. None of these defects is in principle irremediable and in future experiments many of them could be eliminated. For example, arrangements could be made for the dance bands to employ wide-range microphones throughout, and special loudspeaker arrays could be set up enabling a larger number of observers to be tested at one time without introducing unintentional losses at high frequencies. The greatest single instrumental difficulty, and one which caused much loss of time, was the shortage of high-grade transportable recording equipment. Since the commencement of these experiments, better equipment has become available; even so, it will almost certainly be necessary to provide supplementary equalisation if an overall frequency characteristic flat to 15 kc/s is to be a condition of the test.

* It should further be noted that, in such an operation, all observers would have to be medically examined so as either to exclude subjects with hearing abnormalities or, if desired, to include them in representative proportion.

**It is probable that pizzicato effects would be affected significantly by Filter II.

6. REFERENCES

1. B.B.C. Engineering Division Monograph, No. 4. "The Design of a Ribbon Type Pressure-Gradient Microphone for Broadcast Transmission". December 1955.

2. Research Department Report, No. M-026. "The AKG Electrostatic Microphone Type C.12".

3. Research Department Report, No. A-037. "Analysis of Data from Subjective Tests Employing Graded Scales of Observer Preference".

EGF