PREFACE.

This dissertation is submitted for the degree of Ph.D. in Electrical Engineering at the University of Cape Town.

The research for this thesis was carried out as part of the research programme of the Council for Scientific and Industrial Research in their Acoustics Laboratories in Pretoria under the Promotion of Prof. R. Guelke and Supervision of Dr. J.P.A. Lochner.

The investigation covered various aspects of the effects of a single delayed echo on the intelligibility of speech and was extended to cover the case of multiple echoes which is more in line with practical conditions than a single echo.

The application of the results obtained to architectural acoustics and speech reinforcement is discussed.

J.F. Burger

PRETORIA:

11th August, 1955.

CONTENTS

4.4. Effect/....

1. INTRODUCTION

It is well known that the sound energy reaching a listener in an enclosed space, such as a room, consists of two parts, viz. direct sound and reflected or reverberant sound. In the vast majority of cases the total energy received as a result of reflections from walls and ceiling, is at least equal to, and quite frequently much more, than that received directly from the source.

Since the reflected sounds must of necessity travel further than those that reach the listener directly, they constitute in effect a number of echoes. It is a characteristic of the ear that, as far as directional location of the sound source is concerned, only the first pulse of each pulse train is taken into account. The listener, therefore, will judge the direction from which the sound is coming, as being that direction from which he receives the first pulse. In the normal case of a listener in a hall or a room, this will of course be the direct sound.

There is yet another effect, however, which has been qualitatively and quantitatively examined by Haas [1]. This is namely that, provided certain conditions are met, the first pulse arriving at the ear of the listener effectively masks all subsequent echoes of that pulse. Using a single artificial echo, Haas determined by how much the intensity of the echo had to be increased with respect to the intensity of the primary sound to appear to an observer to be just as loud as the primary tone at a given delay time between them. He used speech as testing material; his results are shown in figure 1. It is evident that for a delay time of say, 20 milliseconds, the "masking" of the primary sound is considerable, since for equal loudness the echo must have a power of about ten times

that/...

1) Haas, H.: "Ueber den Einfluss eines Einfachechos auf die Hörsamkeit von Sprache" - Acustica 1 (1951) pp. 49 - 58.

that of the primary sound.

The initial pulse, therefore, serves to indicate the direction from which the sound is coming, and secondly masks the subsequent echoes, which in most cases may be arriving from many different directions in a random manner. Now, although in a well designed auditorium or in a smallish enclosure where the delay times are short, the ear does not notice the echoes as such, their energy is not lost, but contributes very much to the apparent level and intelligibility of the received sound. In fact, an integration of the primary sound and echoes is performed by the ear, analogous to the way in which the eye can integrate a flickering light or light pulses, to obtain an effect of continuity.

The exact way in which the ear integrates the echoes of speech sounds has been a matter of speculation for quite some time. Von Békésy [2] has measured how the loudness of a pure tone increases with time after initiation, and found that it reaches a maximum after about 200 milliseconds, as shown in figure 2. The subsequent decrease is due to fatigue effects.

Munson [3] repeated these measurements, and obtained results which, though differing in some respects, at least confirmed that integration continued up to about 200 milliseconds.

Garner [4] using pure tone pulses of adjustable repetition rate as well as duration, tried more specifically to measure the integration characteristics of the ear, and

came/....

2) V. Békésy, G.: "Zur Theorie des Hörens" - Physik. Zeitschr. March, 1929, p.115.

3) Munson, A.: "The growth of auditory sensation" - J.Acous. Soc.America, July, 1947, p.589.

4) Garner, W.R.: "Auditory Thresholds of short tones as a function of repetition rates" - J. Acous. Soc.America, July, 1947, pp. 600 - 608.

came to the conclusion that integration occurred only over a narrow band of frequencies, provided further that the signal was continuous. This conclusion would seem to exclude speech, which covers a spectrum of a few kilocycles, and consists of sounds essentially transient in character, on the average of roughly 100 milliseconds duration. Yet, integration of speech does most definitely occur, for a simple calculation can show that in a large hall, at a distance of say 60 feet from a speaker, the greater part of the received energy is due to reverberant sound. If the intelligibility of speech were due to the direct sound only, the high ratio of reverberant to direct sound would result in extremely poor articulation. In a well designed hall this is manifestly not the case, leading to the conclusion that considerable integration of the multiplicity of echoes constituting the reverberant sound must take place.

Other workers [5,6], quoted by Haas, measured the integration of a single echo by observing the increase in loudness caused by the echo. The test signal, consisting of speech and music, was presented to the observer, who then had to balance an 800 cycle reference signal for equal loudness. This balance was performed for test signals with and without the echo, and in the majority of cases they found that in the presence of an echo of energy equal to that of the primary sound, the increase in loudness so obtained exceeded 3 phons. This would indicate that the integration process did not consist of simple addition of energies. Haas, however, questions the accuracy of their results. In his case, the observer had to increase the level of the primary sound, after the echo had been switched off, until the primary sound appeared to be just as loud as the composite signal of primary sound

plus/....

5) Aigner, F and Strutt, M.: Zeitschrift für Techn. Physik 14, (1934) p.355.
6) Lüncke, E.: Zeitschrift für Techn. Physik 15 (1934) p. 77.

plus echo had been. His method, therefore, while also being one of equal loudness judgement, differed from that of the previously mentioned workers in that the primary sound itself, instead of a pure tone, was used as reference signal. In the case of delay times not exceeding about 35 milliseconds, he found that without exception the increase in loudness, due to an echo of loudness equal to that of the primary tone, was three phons. This is the result that would be expected if the ear integrated by simple addition of energies.

It has been assumed by some workers [7,8] that the ear integrates speech over a period of 50 milliseconds. Others have variously taken the integration period as 62 milliseconds [9] or only 30 milliseconds [10]. It seems unlikely, however, that the integration period will be a definite time interval over which complete integration takes place, and that outside this interval no integration whatsoever occurs. Rather, one expects that as the time interval increases, integration of the sound energy will gradually decrease, until at some stage a masking effect might even be produced by the echo.

In the design of auditoria, public address systems and the like, the problem of echoes becomes of paramount importance. For the reinforcement of sound, either by natural echoes or artificially delayed amplified sound, it is necessary to know how the contribution of the echo to the intelligibility of speech depends on the delay time. In other

words/.....

7) Petzold, F.: Elementäre Raumakustik, p.8 (Bauweldt)
8) Meyer, E.: "Definition and Diffusion in Rooms", J.Acous. Soc.America 26 (1954) p.634.
9) Zwikker, C.: "Verstaanbaarheid van Luidsprekerinstallaties" - Ingenieur (Den Haag) 44 (1929) p.39.
10) Sawade, S.: "Bemerkungen zu einigen Problemen der Raumakustik" Elektrotechn. Zeit. 71 (1950) p.245.

words, the fraction of the echo energy that is integrated is in all probability a function of time, and it is necessary to know what that function is. Furthermore, to obviate disturbing echoes, it is necessary to know at what intensities, given a certain delay time, an echo becomes audible and finally objectionable. The criterion proposed by Bolt and Doak [11], based on the results obtained by Haas and by themselves, as well as later work by Muncey, Nickson and Dubout [12] shows that the acceptability of echoes depends amongst other things, to a large extent on the reverberation time of the room and that echoes become less noticeable as the reverberation time is increased.

The object of the research herein described was twofold: First, to determine, as a function of delay time, at which intensity an echo became audible and eventually as loud as the primary sound under non-reverberant conditions; and secondly, to determine the law according to which the ear integrates echoes to enhance the intelligibility of speech, and its implications in architectural acoustics.

11) Bolt, R and Doak, P.: "A tentative criterion for the short-time transient response of auditoriums" J.Acous.Soc.America 22 (1950) pp. 507 - 509.

12) Muncey, R.W., Nickson, A and Dubout, P.: "The acceptability of speech and music with a single artificial echo", Acustica 3 (1953) pp. 168 - 173.

2. APPARATUS AND TESTING MATERIAL.

2.1.1. Apparatus:

The apparatus used to produce a single artificial echo is shown diagrammatically in figure 3. An endless loop of magnetic tape is driven at a speed of 15 inches per second by capstan C over two sets of erase, record and playback heads E_1 R_1 P_1 and E_2 R_2 P_2 respectively. Between the record and playback heads of each set of heads, the tape has to pass over pulleys p_1 and p_2. These pulleys were mounted on a carriage which could be moved by a worm gear towards either of the two sets of heads, as indicated by the arrows in the figure.

Signal, applied to the two record heads simultaneously, will be reproduced by the playback heads after a certain delay, determined by the distance the tape has to travel between the record head and its associated playback head. Hence, if the pulleys p_1 and p_2 are exactly in the middle, the delay time between R_1 and P_1 will be the same as the delay time between R_2 and P_2 with the result that there will be no delay between the signal from P_1 and that from P_2. As the carriage carrying p_1 and p_2 is moved from the centre position, the signal from P_1 will lead or lag that from P_2 according to whether the carriage is moved towards P_1 or P_2 respectively. In this manner, delays ranging from 0 to 500 milliseconds could be obtained, with either of the two channels ahead of the other.

2.1.2. Testing Material:

All the test material used was recorded on an Ampex model 408R tape recorder, at a speed of $7\frac{1}{2}$ inches per second. The signal from the Ampex was fed to the record heads of the delay mechanism through a common recording amplifier. From the two playback amplifiers of the delay mechanism, the two

signals/....

The lists were read at a constant speed, one line every seven seconds. Each line consisted of an introductory phrase followed by two test words. The speech level was maintained as constant as possible. A test signal of twelve seconds duration was recorded between lists for calibration purposes.

For the benefit of Afrikaans speaking observers, a set of 17 Afrikaans word lists, similar to the Harvard PB 50 lists, was drawn up, and recorded in exactly the same way. These two recordings were then used in all the tests described hereinafter; about equal numbers of English and Afrikaans speaking observers were used.

3. TESTING PROCEDURE TO DETERMINE "THRESHOLD OF PERCEPTION" AND "EQUAL LOUDNESS" CONTOURS.

3.1. Calibration of apparatus:

To ensure equal output from the two channels, the following method was adopted:

A 1kc warble tone, sweeping through a 100 cycle deviation at a rate of 32 times per second, was fed into the Ampex recorder. The Western Electric condenser microphone was suspended in the anechoic chamber at the spot where the head of a seated observer would normally be. Output from the microphone was fed to the level recorder, which fitted with a 10 db potentiometer, was capable of 0.1 db resolution. Hence, by switching on one channel at a time, it was possible to equalise the levels of the two channels by means of the stepped attenuators, the smallest steps of which were also 0.1 db.

In order to be able to correct for drift in the gain of any of the electronic apparatus, a calibration signal was recorded, as previously mentioned, in the intervals between word lists. This signal made it possible to ensure that the output of the tape recorder was maintained at a fixed level, simply by observing the output meter on the Ampex between lists. By connecting the Ballantine valve voltmeter across the output of the Leak amplifiers, the absolute level of the signal fed to the loudspeakers could be measured. An accuracy better than 0.2 db could be attained in this setting.

It must be pointed out at this stage, however, that although the two channels could be adjusted within say 0.2 db of each other, this did not necessarily imply that both would yield the same percentage articulation. The reason for this

is/....

is that the test signal was still relatively of narrow bandwidth, centered at 1kc, whereas speech fills a much wider band. Had random noise been used as a test signal, this trouble would not have been eliminated either since then the test signal would have covered a much wider band than the speech. As can be seen from the frequency response curves of figure 6, the two channels are very similar over the speech frequency spectrum, so that using a 1kc warble tone would give, to a first approximation at least, equal intelligibility of speech for equal output of test signal.

3.2. The threshold of perception and equal loudness contours.

3.2.1. Determination of threshold contours:

The level at which a single echo just becomes noticeable depends on two factors - the intensity of the primary tone, and the time delay between primary tone and echo. The tests to be described were performed at two different levels of primary tone, viz. at about 25 db above hearing threshold, and at about 50 db above hearing threshold. While the absolute levels of primary tone were kept constant for all observers, the value, referred to hearing threshold, of course varied from observer to observer, depending on the hearing threshold of the observer. The values of 25 db and 50 db mentioned above are with respect to the average hearing threshold of the five observers used in these tests.

Seated in the anechoic chamber, the observer was presented first with signal from the primary loudspeaker only. The signal consisted of Harvard Articulation Lists, as previously described in section 2.2. With a certain fixed delay between the two channels, the echo loudspeaker was gradually turned up, until the observer indicated, by

pressing/....

pressing a button, that it was clearly audible.

Having heard how the character of the sound alters as the echo increases, the observer was next told that he should judge whether he could hear the echo loudspeaker as a separate source during a given group of signals. Each group consisted of three lines of a Harvard list and the observer had to indicate after each group had been heard, whether he had heard the echo loudspeaker or not. Indication was given by pressing one of two buttons, "yes" or "no". Test sessions usually lasted for about half and hour, after which an adequate rest period was allowed.

For each delay time, the intensity of the echo loudspeaker was varied by an integral number of decibels, in more or less random fashion, after each group of three lines had been presented to the observer. After each group, the observer gave his judgement and to facilitiate interpretation, the results were plotted as shown in figure 7, which is a typical example of the results obtained. As will be seen, each "no" or "yes" judgement was noted by means of a dot on the appropriate line [13], at the level at which the judgement was made. During the course of the test, each setting of echo loudspeaker level was presented to the observer twice. A vertical line was then drawn through the points in such a way that the number of "no" judgements falling to the right of this line was equal to the number of "yes" judgements falling to the left of the line. This line was then taken as indicative of the level at which the echo loudspeaker became audible to the observer as a separate source.

The/....

13) This is a standard method. cf.Churcher, B.G. and King, A.J. "The performance of noise meters in terms of the primary standard" - Journal of the Institution of Electrical Engineers, July 1937, p. 57.

The tests were performed, at each value of delay time chosen, with first the one, and then the other channel leading in time. In this way, inequalities that might have existed between the two channels (due to slight differences between the loudspeakers) tend to cancel, and any possible directional bias on the part of the observer is also cancelled. Figure 8 shows the results obtained with one of the observers. The circles indicate that the left hand side channel was leading, and the dots that the right hand side channel was leading.

Five observers were used for these tests, which were carried out with delay times ranging from zero to 120 milliseconds, and in figure 9 the mean results of each observer are shown, and in figure 11 the best curve through the mean of these points, for a primary tone having a level of about 25 db ré hearing threshold. The tests were repeated with the same five observers, with a primary tone level 25 db higher, i.e. at about 50 db above hearing threshold. The results are shown in figures 10 and 11.

3.2.2. Determination of equal loudness contours:

The technique employed in this series of tests was substantially the same as that described in the previous section, except that in this case, of course, the observer did not indicate audibility of the echo loudspeaker, but rather which of the two channels sounded the louder. "Equal" judgements were not permitted.

As before, the tests were carried out at two different levels of primary tone, viz. 25 and 50 db above average hearing threshold. The results, being the mean of the same five observers that carried out the threshold tests of section 3.2.1 are shown in figures 12 and 14 for a primary tone level of

25 db/....

25 db, and in figures 13 and 14 for a level of 50 db.

3.3. Estimate of accuracy of results.

3.3.1. The threshold contours:

To estimate the accuracy of the contour obtained with a primary sound of 25 db above hearing threshold, the points obtained by the five observers at a delay time of 80 milliseconds were considered. Reference to figure 9 shows that the spread at 80 milliseconds is representative of the spread obtained at other delay times, and if anything, slightly greater than average. Calculation shows that the standard deviation [14] at 80 milliseconds is about 2 db.

In the case of the 50 db curve (figure 10) if the points obtained at a delay of 20 milliseconds are considered to be representative, if somewhat pessimistic, of the spread obtained for this curve, the standard deviation works out at 2 db.

3.3.2. The equal loudness contours:

In the case of the 25 db contour, the points obtained at a delay of 50 milliseconds (see figure 12) were used to calculate the standard deviation (about 1 db).

In the case of the 50 db contour, analysis of the points obtained at 50 millisecond delay (figure 13) shows the standard deviation to be about 2 db.

The standard deviations found above are of the order obtained with subjective tests of this sort [15]. At the

shorter/....

14) Mounsey, I: Introduction to statistical calculations p. 139.

15) Richardson, E.G.: Technical Aspects of Sound (Elsevier) p. 162.

shorter delay times, i.e. 20 milliseconds and less, the accuracy is probably a bit less, in view of the larger spread and greater deviation of mean points from a smooth curve, that have been obtained here.

3.4. Discussion of the results obtained.

3.4.1. Threshold of perception contours:

Reference to figure 11 shows that the level of the echo ré the primary sound is not the same function of delay time for both the values of primary sound intensity considered. The general appearance of the curves is similar, though, and it appears that, for a given delay time, the ratio of primary sound to echo must increase if the level of the primary sound increases, if the echo has to remain inaudible.

3.4.2. Equal loudness contours:

Again, the level of the primary sound appears to have some effect on the shape of the curve obtained. Haas did not find this effect, and it is possible that the presence of vestigial echoes in his tests may account for this. Much more interesting, however, is the fact that the curve obtained by the present writer differs somewhat from his, as figure 15 will show. This figure shows the results obtained by Haas, with the 50 db contour of figure 14 superimposed (dotted line). It is possible that these differences are due merely to inaccuracies of measurement or judgement. On the other hand, it is suggested that the vestigial echoes, that have previously been mentioned as possibly being present in Haas's tests, may account for the differences. Another possible source of error might be the fact that his equipment

did/....

did not allow the echo signal to be presented first through the one, then through the other channel. The results obtained can be measurably affected by differences between the two channels in high-frequency response, such as may be occasioned by slight differences in otherwise identical loudspeakers, or slight misalignment of one of the playback heads.

In figure 16 is shown the threshold of perception and equal loudness contours obtained with a 50 db primary sound. The curious fact emerges that for delays of less than 2.5 milliseconds, the threshold contour lies higher than the equal loudness contour. The explanation here is that the observer was told, for determination of the threshold contour, to press the "yes" button only when he could definitely hear sound coming from the echo loudspeaker, as a separate source. Now, at fairly longish delays, this is easy since a definite echo can be heard when the echo loudspeaker reaches a certain level. At shorter delays, however, no distinct echo can be heard. As the echo loudspeaker increases in intensity, a change in the quality of the sound is first noticed. The initiated listener will realize that this is due to the echo loudspeaker, and will therefore, strictly speaking, be aware of its presence. Still, at this point, had the observer not previously known of the presence of the echo loudspeaker, he would not yet have become aware of its presence. Hence the instruction to the observer to signal awareness only when he could definitely hear signal coming from the echo loudspeaker. Consequently at very short delays the observer sensed that the two loudspeakers were on, and even judged them to be equally loud, long before the echo loudspeaker could, as it were, override the primary loudspeaker

and/....

and force its attention on the observer. Observers also experienced much greater difficulty in the threshold judgement for delay times up to 10 milliseconds than for the longer delays.

Finally, mention must be made of a curious phenomenon observed in the course of these tests. As has already been pointed out, the delay mechanism allowed the echo to pass through zero from the one to the other channel. When this was performed slowly, with both channels at the same level, it gave the impression to observers in the anechoic chamber that the apparent sound source moved from one side to the other, passing close to the head, somewhat higher and to the front of the observer. With no delay, the sound source seemed to be, roughly, in the same position relative to the observer as the live speaker had been relative to the microphone when the recording of the testing material had been made. This was observed for the case where the two loudspeakers were connected in phase. When they were reconnected to be in antiphase the sound source seemed to travel again from one side to the other as the delay was moved through the zero position, but this time the apparent sound source, at zero delay time, was located seemingly inside the head of the observer, slightly toward the back, but at the level of the ears. This effect was obtained at every point in the anechoic chamber equidistant from the two loudspeakers. When a slight delay was introduced, moving to a point closer to the echo loudspeaker and further from the primary loudspeaker, again brought on the sensation that the sound source was inside the head of the observer, when a point was reached where both signals again arrived simultaneously.

The condition where identical signals arrive simultaneously but in antiphase from two different directions

at/....

4. INTEGRATION CHARACTERISTICS OF THE EAR.

4.1.1. General considerations:

To measure the integrating characteristics of the ear, Harvard PB 50 articulation test lists were used, as in the previous tests. As is well known, the articulation score obtained by an observer under a given set of conditions can be distressingly variable, and the variation from observer to observer much greater still. Clearly, therefore, to obtain reasonably reliable results with articulation tests, the following conditions, necessary but not sufficient, have to be met.

(1) A great number of tests must be performed, in order to obtain a reliable average.

(2) Tests conducted with a given observer must, for the sake of consistency, be carried out in as short a time as possible. It is not advisable, for example, to carry out one half of a test on one day, and the other half the next day.

(3) Reasonable care should be exercised in the choice of an observer. A person who is incapable of sustained albeit not too severe concentration, will yield disappointing results. Unfortunately, this inability is usually only detected after at least one test has been performed.

4.1.2. Principle:

As is known, the percentage articulation obtained by a given observer is, all other things being constant, a function of the level at which the words are heard. In figure 15 is shown a curve of percentage articulation vs. level in db ré hearing threshold. The curve shown was obtained

during/....

articulation vs. power output of one of the channels would likewise have shown a shift of 3 db, since again a doubling of the sound energy at the ears of the observer would have occurred.

It now remained to be seen if this method would work in practice, and what degree of accuracy could be obtained with it. If successful, the method could be used to measure the integrating properties of the ear, by simply finding, for a given delay, what the horizontal shift of the curve would be as compared to one obtained with zero delay between the two channels.

The procedure followed in the preliminary tests to check the accuracy of the method, and in the subsequent tests with delayed echoes, is described in the following section.

4.1.3. Description of preliminary tests to check the accuracy of the method:

The apparatus was set up exactly as for the equal loudness and threshold of perception tests described in previous sections. The calibration was performed as set out in section 3.1. Test sessions were started off by playing to the observer, seated in the anechoic chamber, a word list through one of the channels. Beginning at a low level of only a few db above hearing threshold, the same list was played over about four times, the level being increased by five db after each playing. This served to accustom the observer to the tests, and also showed roughly at which level he attained a 50% articulation score. This level varied from observer to observer, depending amongst other things, also on their individual hearing thresholds.

Having ascertained where the observer would score 50%, the test proper was started by letting the observer

write three complete lists, presented to him through one channel only, at a level about 10 db lower than that required for 50% articulation. This channel will subsequently be referred to as "the first channel". Then, the level was increased by 4 db by means of the Daven attenuator and the same three lists again played to the observer. This process was repeated twice more, the level being increased by 4 db for each repetition. Such a session took about 45 minutes, and after its conclusion the observer marked the lists he had written. This part of the test will subsequently be referred to as Test 1.

The results were plotted as shown in figure 18. The mean of the three scores obtained at each level was found, and the best straight line drawn through the four points thus obtained. The level at which a score of 50% was obtained was then obtained from this line. It will be noticed that the levels at which the lists were presented, were so chosen that the scores obtained straddled the 50% articulation point. Figure 18 is a typical example of the results obtained.

Next, the two attenuator-amplifier-speaker channels were connected in parallel to the output terminal of the left hand side channel of the delay mechanism. Effectively therefore, the left hand side and right hand side speaker channels were in parallel since both received the same signal from the left hand side channel of the delay mechanism. By means of the Daven attenuators the two channels could be adjusted for equal output, using the Western Electric condenser microphone and Brüel and Kjaer level recorder to measure the loudspeaker outputs, as described in a previous section. By using the Ballantine valve voltmeter to measure the voltage fed to the speaker used in the test just described, it could be reset to exactly the same levels as in the test

where/....

where the second channel was not connected in parallel. This was necessary since connecting the two attenuators in parallel to the delay mechanism loaded the latter so that a drop in output of approximately 2 db occurred.

Three further lists were now played to the observer. As in the first part of the test, these lists were repeated at increasing levels, until the observer had written the group of three lists at four levels. As before, these levels were spaced three or four db apart, and were so chosen that they straddled the level required for 50% articulation. During the course of the test, the two loudspeaker channels were of course kept in step by varying both Daven attenuators simultaneously. The levels noted were, however, those of the first channel. They referred (being in db) to the same arbitrary reference level of electrical output from the Leak power amplifier, as was used as reference in the first part of the test. The results of this second test, subsequently referred to as Test 2, were plotted exactly as before (figure 18).

Finally, the apparatus was reconnected as for Test 1. This time the first channel was muted, after the channels had been readjusted to obtain equality. The second channel was now used, and the observer was given yet three more word lists at four levels, exactly as before in the first part of the test when the other channel was used. Again, the results were plotted as before and the best straight line drawn through the mean points obtained. The levels in this part of the test were referred to a reference level which would, with this second channel, give the same acoustic output (as measured with the condenser microphone as before) as the reference level, used in test 1 and test 2, would give with the first channel, and this test will be referred to in

future/....

future as test 3.

An illustrative example might help to clarify the measuring procedure just described.

Test 1

50 millivolts across speaker 1 voice coil assumed to give a level 31 db above hearing threshold.

Channel 1 attenuator setting for 50 millivolts across speaker 1 voice coil:- 5.7 db

Channel 2 attenuator setting for equal sound pressure (measured with condenser microphone):- 6.3 db

Voltage across voice coil of channel 2 speaker with attenuator set to 6.3 db:- 51 millivolts

With channel 2 muted, articulation lists are presented to the observer at the following settings of channel 1 attenuator (db)

27.7 24.7 21.7 18.7

Hence, the corresponding respective levels ré the assumed arbitrary hearing thresholds are

9.0 12.0 15.0 18.0 db

Test 2

Channels 1 and 2 connected together to one of the output amplifiers of the delay mechanism.

Channel 1 attenuator setting for 50 millivolts across speaker 1 voice coil: 4.6 db.

Channel 2 attenuator setting for equal sound pressure (measured with condenser microphone) 5.6 db.

Articulation lists are presented at the following attenuator settings:

Channel 1 attenuator: 29.6 26.6 23.6 20.6

Channel 2 attenuator: 30.6 27.6 24.6 21.6

Hence the level of <u>each</u> channel ré the same hearing threshold as assumed for test 1, is 6.0 9.0 12.0 15.0 db respectively.

Test 3/....

Test 3

Channels reconnected as for test 1.
Channel 2 attenuator setting for 51 millivolts across speaker 2 voice coil: 6.3 db.
Channel 1 muted, lists presented with channel 2 attenuator settings of 28.3 25.3 22.3 19.3 db. corresponding to levels ré the previous assumed threshold of 9.0 12.0 15.0 18.0 db respectively.

These are not actual figures, but were chosen merely for the purpose of illustration.

To resume. The levels at which the observer scored 50% were found from the best straight lines drawn through the points obtained. The mean of the levels required for 50% articulation in tests 1 and 3 should theoretically be 3 db higher than the 50% level of test 2 since in the latter test the power was twice as much as in tests 1 and 3.

As is evident from the foregoing description, the tests were performed to the following pattern: One channel, both channels together, finally the other channel. The reason for adopting this procedure was threefold.

(1) Some observers were found to attain consistently higher scores when the sound reached them from one particular direction than when it came to them from the other direction. Whatever the reason for this effect, errors which would have arisen had the single channel test been confined to one where the sound came from only one direction, tend to cancel. Since in the case where both channels are on the sound reaches the observer from both directions, the result of this test can only be compared to the mean of the results obtained with sound coming first from one, then from the other side.

(2) Some/....

(2) Some observers continued "learning" as the tests continued. This means that the level required for 50% articulation under otherwise constant conditions, would decrease with each successive test. Of course, there is a limit to this learning process, but provided it is a linear function of the number of tests performed for the three tests under consideration, the effect can be nullified by placing the "two channel" test in the middle. For example, if the 50% articulation level decreases by n db for each successive test, the second test would yield an answer that was n db low. The third test would give an answer 2n db low. And the mean of the first and third tests would be n db low, exactly the same as the second test, resulting in complete cancellation of this error. Even where the "learning" is not a linear function of number of tests performed, this method still minimizes errors arising from this cause.

(3) Possible inequalities that might exist between the two channels because of slight differences between the two loudspeakers are to a large extent cancelled. It must be remembered that the warble tone test signal did not cover the whole speech frequency spectrum band, and therefore equal warble tone output did not of necessity imply equal intelligibility.

4.1.4. Results of preliminary tests with zero delay:

The results obtained are given in the following table:

26/....

Observer	Level (in db ré an arbitrary reference) required for 50% articulation.				
	Channel 1	Channel 2	Mean	Both Channels	Difference
J.S.	15.1	13.9	14.5	11.0	3.5
P.L.B.	14.2	16.6	15.4	12.3	3.1
P.B.	15.4	15.0	15.2	12.2	3.0
J.S.	15.9	14.8	15.4	12.4	3.0
					Mean value 3.1

The results tabulated show a standard deviation of 0.2 db. The method is therefore probably of sufficient accuracy to warrant an attempt to use it in determining the integrating characteristics of the ear for a single delayed echo.

4.2. Tests with a single delayed echo of the same intensity as the primary sound.

4.2.1. Method:

Fundamentally, the method is the same as that used in the preliminary tests; the procedure adopted was as follows: calibration and adjustment of the equipment proceeded as set out in section 3.1. A single word list was played to the observer at increasing levels, to obtain a rough idea of the signal strength required for 50% articulation. Once this had been obtained, the following sequence was followed.

Test A. With a certain delay between the channels (adjusted for equal output as described in section 3.1) three word lists were given to the observer at a level chosen to yield a percentage articulation of about 20%. The level of both channels was then increased by 3 db and the same three lists

repeated/...

repeated. This was repeated for two more levels, each 3db higher than the previous one, giving articulation percentages that would straddle the 50% articulation mark. In short, this test was a repetition of test 1 described in detail in section 4.1.3 except that the second channel was also switched on, and delayed by a definite time interval with respect to channel 1. The results were plotted exactly as before, and the best straight line drawn through the four mean values of articulation percentage obtained.

Test B. For this test, the two channels were again connected in parallel to one of the output amplifiers of the delay mechanism. This test was an exact repetition of test 2 of section 4.1.3. The results were plotted in exactly the same way.

Test C proceeded in the same way as test A except that the second channel now supplied the primary sound and the first channel the echo. This was achieved by adjusting the delay on the delay mechanism to minus the value it had been set to for test A of this section. In other words, the pulleys p_1 and p_2 were set to the opposite side of the zero position (see figure 3) at a distance from the zero position equal to that employed in test A. The results were plotted as for test A.

Briefly, therefore, the sequence was: Echo (channel 2 behind channel 1), no delay. Echo (channel 1 behind channel 2). Once more, the example given on page 23 will serve to illustrate the relations between the various levels, except of course for the fact that now both channels were adjusted for equal output during all three parts of the experiment, and not only during the second test.

As before, the levels (ré a constant arbitrary threshold) required for 50% articulation were found. The mean of the 50% level of test A and the 50% level of test C was obtained, and compared with the 50% level of test B. Since test B represented in effect perfect integration, it follows that less than perfect integration would result in a displacement of the curves of tests A and C towards the right. In other words, tests A and C would require a higher level for 50% articulation than test B. If the mean of the 50% levels of tests A and C were 3 db higher than the 50% level of test B, this would imply that neither integration nor masking had taken place - in other words, the echo had had no effect.

One more point about the method employed here perhaps needs additional elucidation. It is clear that the extent to which an echo is subjectively integrated with the primary signal of equal intensity can be determined in two ways - by comparison with a single signal or by comparison with two perfectly integrated signals with equal intensities, i.e. two equal signals with zero delay between them. Stated otherwise, one can determine by how much the presence of the echo improves the effective level above that obtained with the primary tone alone, or one can determine by how much the echo falls short of being perfectly integrated.

If the former comparison had been made, two reference tests would have been necessary - one with channel 1 alone, and one with channel 2 alone, resulting in a total of four tests. If, however, the two reference tests are made simultaneously, i.e. both channels on simultaneously, one test suffices as reference, but this reference is then of course the case where perfect integration takes place.

4.2.2. Results:

The measurements described in the previous section were performed for delay times ranging from zero to 120 milliseconds. The results of these measurements are shown in figure 19, which depicts, as a function of delay time, by how many db the equivalent signal level was increased by the presence of the echo. Or stated in other words, by how much, as far as the ear was concerned, the useful signal had been increased by the echo.

Figure 19 may be redrawn to show what fraction of the echo had been integrated by the ear. For example, since the primary tone and echo were always of equal power, an increase in effective level of 2 db means that 0.58 of the energy of the echo has been integrated by the ear.

Figure 20 shows the relation of fraction of echo energy integrated with a primary signal of equal intensity vs. delay time between the two signals in milliseconds.

4.3. Effect of an echo of -5 db ré primary tone.

4.3.1. Method:

To investigate how the integration characteristic of the ear was affected by the level of the echo, the methods used in determining the 0 db echo integration curve were adopted for the case of an echo of level -5 db ré the primary sound.

Setting up of the apparatus proceeded as in the case of the 0 db echo. The test sequence and procedure also remained the same, except that now, in tests A and C (refer to section 4.2.1.) the echo channel was adjusted to a level of -5 db with respect to the leading channel. Test B remained unaltered with both channels set to equal levels.

As/...

As the reference test B represented perfect integration of two equal signals, the mean 50% level obtained from tests A and C should be displaced by 1.8 db if perfect integration of the -5 db echo took place. This is, of course, because addition of energies at levels of 0 and -5 db results in an increase of 1.2 db. As before, test B was done with both channels equal because this took care of both channels (and both ears of the observer) simultaneously, thereby reducing by one the number of tests that would otherwise have been necessary.

4.3.2. Results:

The curve of figure 21 shows, for delays between zero and 100 milliseconds, how the energy of a -5 db echo is integrated by the ear. To be more precise, it shows, as did figure 19 how the effective signal level is affected by the presence of the echo.

4.4. Effect of an echo of level +5 db ré the primary tone.

4.4.1. Method:

The method used was identical with that used for the -5 db echo, except that in tests A and C the echo channel was adjusted to a level of +5 db ré the leading channel.

4.4.2. Results:

The results of these measurements are shown in figure 22 for delay times between 0 and 105 milliseconds.

4.5. Dependence of the integration characteristic of the ear on the level of the primary signal.

4.5.1. General considerations:

To determine how the absolute level of the primary sound and the echo affects the integration characteristic of the ear, it was necessary to perform at least the tests with

an/...

an echo of equal loudness at a considerably higher level than before. Since in the tests described in section 4.2.1. the absolute levels at which the tests were performed were determined by the levels at which the observers obtained 50% articulation, these levels were rather low, being of the order of 25 db ré 0.0002 dynes/cm^2.

In order to be able to increase the levels without also increasing the percentage articulation, it was necessary to introduce some artificial masking. This was done by using random noise as an interfering signal.

4.5.2. Method:

The masking noise came from a loudspeaker suspended in the anechoic chamber at a distance of about 8 feet from behind the seated observer, and directed at him. The noise level produced by this loudspeaker was adjusted to give 50% articulation scores at a level about 25 db higher than was the case in the previous tests.

The tests described in detail in section 4.2.1. were then repeated, in the presence of the masking noise. Figure 23 shows the results obtained (dots). The solid line is that of figure 19 and is shown for reference purposes.

4.6. The integration of multiple echoes.

4.6.1. General remarks:

In figure 20 it was shown how the fraction of the single echo integrated by the ear varied with delay time. In practice, however, single echoes are comparatively rare, and it is therefore of interest to know if this integration characteristic holds also for multiple echoes. A series of experiments was therefore performed, in which a primary tone and three echoes were used.

4.6.2. Apparatus:

The mechanism used to obtain the four signals required, consisted of a "Pamphonic" delayed sound reproduction apparatus. Essentially, this machine comprises a recording amplifier, and the usual erase and bias high-frequency sources normally associated with tape recorders; one recording, one erase, and four playback heads; and four independent playback amplifiers. Instead of on tape, this machine records signals fed into it on a rotating disc which has been given a magnetic coating. The six heads are spaced around the periphery of the rotating disc, and are arranged in the following sequence: record head, four playback heads, erase head.

Continuous operation therefore occurs, in which a signal is carried by the rotating disc from the record head, past the four playback heads in succession, and is finally erased. By altering the spacing between the four playback heads, different delay times may be obtained ranging from roughly 25 milliseconds to about half a second. A separate gain control in each playback channel enables the level of each channel to be adjusted quite independently from that of any of the others.

The Ampex tape recorder was used, as before, to supply the testing material, which was fed to the "Pamphonic" delay mechanism through a low-pass filter cutting off at 4 kcs. This filter was found to be necessary, as otherwise the high frequency responses of the four playback channels were too dissimilar.

From the delay mechanism, the four signals went to four output amplifiers and thence to four Goodmans Axiom 80 loudspeakers, mounted on 24" x 24" baffles in the anechoic chamber. The height of the loudspeakers above the floor was

chosen/...

chosen to be equal to the height above the floor of the ears of a seated observer. Figure 24 shows the arrangement in the anechoic chamber, and the block diagram of figure 25 shows how the equipment was set up.

As before, a warble tone oscillator and Ballantine valve voltmeter were available for calibration, and a Western Electric condenser microphone and Brüel and Kjaer high speed level recorder were used to obtain the frequency response of the whole chain of equipment. The microphone was hung where the head of a seated observer would be, and the response of each channel measured, with the low pass filter set to cut off at 4 kilocycles. The curves of figure 26 show the overall response of the four channels.

The channels were numbered from 1 to 4 in the sequence in which their playback heads were arranged. Taking channel 1 as reference, the delays of the other channels were, with respect to channel 1:

channel 2	23 milliseconds
channel 3	52 milliseconds
channel 4	78 milliseconds

Placement of the speakers connected to these channels is shown in figure 24. The baffles were orientated so that the observer sat on the axis of each of the speakers.

4.6.3. Method:

Reference to figure 26 shows that the overall frequency response of the equipment differed slightly from channel to channel. These differences, due to the loud-speakers, were not serious from the point of view of frequency response alone, but as far as intelligibility went, subsequent tests showed that they did have a noticeable effect. It was

therefore/...

therefore decided, since channels could no longer be changed round as was done in the case of a single echo, not to use the condenser microphone in an attempt to adjust the four channels to equality before each test, but rather to assume them all equal to start off with, and then apply suitable corrections after all the tests had been performed.

Preliminary tests with articulation lists brought to light that channel 2 yielded higher articulation scores than channels 1 and 3 (which gave equal scores). The higher scores of channel 2 corresponded roughly to a level increase of 1 db. Channel 4, on the other hand, gave lower scores than channels 1 or 3, corresponding to a level of about 1 db lower than that of these channels. In all cases, level settings of the 4 channels were made by using a warble tone and measuring the voltage appearing across the voice coils of the loudspeakers.

The method used in determining how the ear integrates the three echoes will now be outlined, and subsequently discussed in greater detail. The observer, seated in the anechoic chamber was presented with articulation tests, played through one of the four channels. After marking his results, he again heard lists, played through another of the four channels. In the next part of the test, all four channels were switched on. And finally, the last two tests proceeded as the first two, but with the remaining two channels. Five tests were therefore performed by each observer; in every case the third test was one in which all the channels were on. The sequence in which the single channels were presented varied from observer to observer, since otherwise the final assessment of the intelligibility of each individual channel might have been affected by its position in the test sequence.

Working/....

Working on the assumption that all four channels were identical, a common reference level was chosen for them. The level taken as reference was one which was produced by a 1 millivolt warble tone signal across the voice coils of the loudspeakers. The warble tone used was one of 500 cycles, with 100 cycles deviation at a rate of 32 per second. The voltage across the voice coil was measured with the Ballantine valve voltmeter.

Gain settings of the apparatus were performed as follows: The gain control of the output amplifier of the Ampex tape recorder was set to give a certain reference voltage (in point of fact, 154 millivolts) across the output terminals when the calibration signal recorded between articulation lists was played back. Then, the warble tone calibration signal described above was fed into the Ampex through its recording amplifier, and the warble tone voltage adjusted so that the same reference voltage appeared across the output terminals of the Ampex as was measured when the calibration signal recorded between lists was played back.

Next, with the warble tone still being fed through the Ampex tape recorder, the Ballantine valve voltmeter was connected across the voice coil terminals of one of the speakers. The gain control of the relevant playback amplifier of the "Pamphonic" delay equipment was then adjusted until the voltage appearing across the loudspeaker voice coil was of the desired level ré 1 millivolt.

It must be made clear at this stage that all levels referred to in the description of these tests will be with respect to this 1 millivolt warble tone signal across the speaker terminals, unless definitely stated otherwise. Therefore, if it is said, for example, that an observer attained a 50% articulation score at a level of 10 db ré 1 millivolt,

it/....

it means that the observer attained 50% articulation with the gain of the channel used set to such a value, that a warble tone output of 154 millivolt from the Ampex tape recorder produced a warble tone voltage across the loudspeaker voice coil that had a level of 10 db ré 1 millivolt.

To resume. An initial test was performed with each observer, in which one test list was played to him through one of the channels at levels of 5, 10, 15 and 20 db ré 1 millivolt. This served to condition the observer to some extent, and indicated more or less at which level he would score 50%. The test proper was then started, and consisted of five parts.

Part 1. Two word lists were played through one of the channels, to the observer seated in the anechoic chamber. They were played at a level about 6 db lower than would yield 50% articulation. Subsequently, the level was increased by 4 db, and the same two lists repeated. All in all, this was done four times, so that the last time the two lists were played to the observer, this was done at a level of +6 db ré the one giving a 50% articulation score. The observer then marked his lists, and the scores were plotted as a function of level precisely as was done in the case of the experiments with a single echo. The mean score attained at each level was found, and the best straight line drawn through these four points. During this whole test, the other three channels were of course muted.

Part 2 of the test was in all respects similar to part 1, except that another of the four channels was used.

Part 3. In this part of the test, all four channels were used. It was anticipated that the integration of the ear would result in an effective level about 4 db higher than that produced by each of the channels individually, consequently each channel

was/....

was adjusted to a level about 9 or 10 db lower than that required for the channel to yield, in the absence of the other channels, an articulation percentage of 50%; otherwise of course, the first lists played would already have yielded very nearly 50% when all four channels were switched on.

As has been previously stated, there was reason to doubt that the four loudspeakers were identical. The four channels were therefore, in the majority of cases, adjusted slightly to try to compensate for these differences, so that an approximately equal contribution - disregarding the effect of delay time - would be made by each. Their relative levels, taking channel 1 as reference, were therefore adjusted as follows: channel 1: 0 db; channel 2: -1 db; channel 3: 0 db; and channel 4: + 1 db.

Three lists were then played to the observer, and repeated at levels increasing by 4 db with each repetition, until each group of lists had been played four times. The scores were again plotted as a function of level, and the mean score at each level was found. The best straight line through these four mean points was drawn.

Parts 4 and 5. These were exact repetitions of parts 1 and 2, except that the remaining two channels were used.

It will have been noted that in parts 1, 2, 4 and 5 only two lists were used at each level instead of three as in earlier experiments. This was unfortunately necessary, since otherwise a complete test would have taken too long, making it impossible to complete it in one day. Some accuracy was probably lost hereby; test 3 was, however, performed with three lists, since the results of test 3 gave the degree of integration that had occurred and this one test therefore required a higher degree of accuracy than the four others.

4.6.4. Interpretation of results:

The way in which the results were interpreted will best be explained by means of an illustrative example. The figures used in this example, are typical of the values obtained. The results of a certain observer (F) were as follows: The levels at which articulation scores of 50% were attained (read from the best straight line drawn through the points obtained) are:

Test 1	Channel 3:	17.7 db	
Test 2	Channel 1:	15.2 db	
Test 3	All four channels simultaneously:	10.8 db	Levels refer to 1 millivolt across loudspeaker voice coil.
Test 4	Channel 2:	12.6 db	
Test 5	Channel 4:	14.4 db	

The relative levels of the four channels during test 3 were: channel 1: 0 db; channel 2: -1 db; channel 3: 0 db; channel 4: 0 db.

Consequently during test 3 each channel, in the absence of all the others, would have given a 50% score at the following levels: channel 3: 17.7 db; channel 1: 15.2 db; channel 2: 13.6 db; channel 4: 14.4 db. If one finds the mean of these four levels, then the average level at which one channel alone would have given a 50% score is $\frac{60.9}{4}$ or 15.2 db. But with all four channels on, the 50% level was found to be 10.8 db hence the integration of the four signals has resulted in a level increase of 4.4 db above the level of one channel.

To determine what the differences, if any, between the four channels were, deviations of the results of each individual channel from the mean result of the four channels were found for each observer. That is to say, for each observer the mean of the four 50% levels, obtained with the

four/....

four channels separately, was found. The difference between this mean 50% level and the 50% level of each individual channel was noted for each observer. This was done for 7 observers, and the mean deviation from the mean found for each channel. The mean deviations for each channel were: channel 1: 0 db; channel 2: +1.75 db; channel 3: +0.25 db; channel 4: -2.0 db.

It was now possible to calculate what the increase in effective level would have been if the results obtained for a single echo of 0 db ré the primary sound, still applied in the case of three echoes of slightly differing intensity. In the example given above, test 3 was conducted at the relative level settings shown. Correcting now for the deviations of the four channels given in the previous paragraph, the relative levels of the four channels in test 3 now become: channel 1: 0 db; channel 2: +0.75 db; channel 3: +0.25 db; channel 4: -2.0 db. and the relative energies, taking channel 1 as reference become:- channel 1: 1.0; channel 2: 1.19; channel 3: 1.06; channel 4: 0.63.

The delay times of these four channels were as follows: channel 1: 0 milliseconds; channel 2: 23 milliseconds; channel 3: 52 milliseconds; channel 4: 78 milliseconds and from figure 20 the fractions of the energy integrated by the ear may be read as 1.00 1.00 0.50 and 0.10 respectively for these delay times. The energy contribution of each channel therefore became: channel 1: 1.00; channel 2: 1.19; channel 3: 0.50 x 1.06 = 0.53; channel 4: 0.10 x 0.63 = 0.06 , giving a total of 2.78 or 4.4 db ré the mean level.

Hence, in this example, calculation based on the results obtained with a single echo predicted an integration equivalent to a 4.4 db level increase, and measurement gave an answer of 4.4. db.

4.6.5. Results:

The measurements and calculations described in the previous section were carried out for each observer. The results are given in the following table which shows the measured and calculated equivalent increase in level due to integration.

Table 2

Observer	Increase in level as result of integration (db)		
	Calc.	Meas.	Error (db)
M	4.3	2.8	1.5
V	3.5	3.2	0.3
dL	4.3	3.1	1.2
F	4.4	4.4	0.0
R	4.8	4.6	0.2
vA	3.8	3.1	0.7
dP	4.3	4.7	- 0.4
			Mean 0.5 db

4.7. Discussion of results.

4.7.1. The results obtained with a single 0 db echo:

As the curve of figure 19 shows, a single echo is integrated perfectly by the ear up to a delay time of about 30 milliseconds, i.e. the energy of an echo occurring between zero and 30 milliseconds after the primary tone, is simply added to the energy of the primary tone. For delay times longer than 30 milliseconds, the increase in effective level due to the contribution of the echo is less than 3 db, and decreases until at a delay time of 90 milliseconds the effective level of primary sound plus echo is equal to the

level/....

level of the primary sound alone. This means that the ear integrates less and less of the echo energy as the delay time increases, until at 90 milliseconds the echo has, as far as intelligibility is concerned, no effect at all. This condition remains practically unchanged from 90 to 120 milliseconds, which was the longest delay time investigated. Taking into consideration the accuracy with which the apparatus could be set up and the deviations of the points obtained from a smooth curve, it is estimated that an accuracy of the order of 0.7 db was attained. Two or three observers gave results in parts A, B or C of the test through which no straight line could be drawn by anyone with a conscience. Their results were therefore not taken into account. Otherwise, no special selection of observers occurred. Any available staff member of the C.S.I.R. was used, provided he or she had normal hearing. Not one had had previous experience of this type of work.

4.7.2. Results obtained with a single -5 db echo:

Reference to figure 21 shows that the curve obtained with this value of echo is somewhat similar to that obtained with the 0 db echo. Perfect integration appears to continue somewhat longer, however, taking place according to the results obtained, for delay times up to 40 milliseconds. After 40 milliseconds, less and less of the echo is integrated, and as in the case of the 0 db echo, integration ceases at about 90 milliseconds. From 90 milliseconds to 105 milliseconds (the longest delay time investigated) the echo has no apparent effect on the intelligibility of speech. The accuracy is estimated as being the same as in the case of the 0 db echo; but, if the results obtained with the -5 db echo are used to evaluate the fraction of the echo that was integrated, the

accuracy obtained will be rather low, since the maximum effect the echo could have was only 1.2 db as compared to 3 db in the case of the 0 db echo.

4.7.3. Results obtained with a +5 db echo:

Figure 22 shows the results obtained; the integration decreases fairly rapidly as the delay increases, and nowhere is a flat portion of the curve in evidence as in the cases of 0 db and -5 db echoes. It would appear that the echo, as a result of its greater relative intensity, gives rise to a much greater effect of confusion than was the case for the weaker echoes, and that as a result the integration is impaired.

At a time delay of about 55 milliseconds a queer phenomenon can be observed. It appears as if the curve would follow one of two fairly well defined sets of points. As there was no reason to doubt the correctness - within 0.7 db as before - of any of the points individually, it was decided to seek the reason elsewhere why two whole series of points should show such deviation.

After some consideration, a possible reason for this divergence presented itself: Because of its higher intensity, some observers apparently could concentrate on the echo rather than on the primary tone when the time delay between them reached the 55 milliseconds mark. Had the echo alone been present, the curve obtained would of course have remained at the +5 db level. However, in concentrating on the echo, the observers were probably disturbed by the masking effect of the primary tone, with the result that the curve reverts only gradually to the +5 db level as the delay between primary tone and echo is increased. As far as the dotted portion of the curve is concerned, this appears to be a continuation of the curve obtained for delay times shorter than 55 milliseconds.

The observer continues to concentrate on the primary sound, and the masking due to the intense echo becomes more and more severe, until at delay times of the order of 80 milliseconds a maximum masking effect is obtained. For longer delay times, the echo begins to dominate, and the curve too, seems to revert to the +5 db level. Apparently, therefore, at long delays the ear concentrates on the stronger signal irrespective of whether it be primary sound or echo. In the case where the echo is the stronger, the primary sound is still of importance at short delays but as the delay increases, a point is reached where attention begins to focus on the echo rather than on the primary sound. Seemingly the delay time where this occurs depends to some extent on the observer, and the variations between observers gave rise to the double curve at delays in excess of 55 milliseconds.

4.7.4. Results obtained in the presence of masking noise:

This experiment was performed in an attempt to determine if the integration of the ear was at all affected by the absolute level of the primary sound and echo. As will be noticed from figure 23 the points obtained show a very wide scatter; the dotted curve represents about the best smooth curve that can be drawn through the points; the solid line shows the results obtained in the absence of masking noise. It would appear that at the higher level, the integration has been somehwat reduced, although the uncertainty of measurement that is evident from the large scatter obtained makes any definite conclusion, based on these results alone, virtually impossible. In view of the results obtained by other workers at higher sound pressure levels, discussed in section 4.8, it would appear that integration proceeds in much the same fashion at different loudness levels, ranging from about 25 phons to 80 phons.

It/....

It is suspected that the inconclusive results obtained were caused by the manner in which the masking noise was introduced. As has been previously described, the loudspeaker which was fed with random noise hung directly behind, and slightly above, the seated observer, at a distance of about 8 feet from his head. If this loudspeaker is considered to be a point source (a sufficiently accurate assumption for at least the lower frequency components, which have the greatest masking effect) then, since very nearly free-field conditions exist in the anechoic chamber, the sound power per unit area decreases as the square of the distance. Calculation shows that increasing the distance between the speaker and the head of the observer from 8 feet to 9 feet would, under these conditions, result in a decrease of 1 db in the level of the masking noise. As it was nearly inevitable that the observer would alter his position slightly from time to time during the performance of a test, the erratic results obtained might quite possible have been due to this effect.

A much more satisfactory method would have been to use two loudspeakers to produce the masking noise, placing them as close to the two signal loudspeakers as practicable. Unfortunately, the number of observers available at this stage was not very great, and it was decided that they could be more profitably employed in the tests with multiple echoes described in section 4.6. In this connection, it may be explained that an observer was found to possess only a very restricted period of usefulness - restricted, in fact, by the number of articulation lists available. It was found that, after having heard a particular list a few times and corrected their results, some observers showed a regrettable proclivity to remember some of the words. This had the

result/....

result that on hearing the test again, their results could be erratic, depending on how much time had elapsed since they had first heard the test, on how many words they had remembered, on how soon they recognized the list, and similar factors. Erratic results were not infrequently obtained when it was attempted to re-use an observer with a retentive memory, and this practice was therefore abandoned.

4.7.4. Results obtained with multiple echoes:

As table 2 shows, the experimental results obtained in these tests are in fair agreement with the calculations based on the assumption that the ear integrates multiple echoes in the same way as a single echo. Nevertheless, the differences between calculated and measured values show a definite trend to one direction, the mean difference amounts to nearly 0.5 of a decibel. This means that the integration is slightly less than would have been the case had the three echoes been integrated in the same way as a single echo.

It is possible that the error would have been smaller had channel 2 been adjusted to the same level as channel 1 in that part of the tests where all four channels were on simultaneously. As has been shown, channel 2, although an attempt had been made to equalize it, was still 0.75 db higher than channel 1. This could unfortunately only be discovered after the tests had been completed, since the mean of a number of test results was required to assess the contribution of each channel. The tests performed with a single echo seemed to indicate that the longest delay time at which perfect integration still occurred, depended on the level of the echo with respect to the primary tone. For an echo of -5 db this was about 40 milliseconds; for a 0 db echo, 30 milliseconds; and for a +5 db echo no definite break in the curve

is/....

is detectable and it would appear that perfect integration occurs only when no delay exists between the two signals.

Channel 2 gave rise to the first echo, at a time of 23 milliseconds after the primary sound due to channel 1. Since for a 0 db echo perfect integration is apparently still obtained at a delay time of 23 milliseconds, the calculations were performed on the assumption that the contribution from channel 2 was perfectly integrated. In view of the fact, later discovered, that channel 2 was still at a level of +0.75 db with respect to channel 1 when the tests were carried out, it seems possible that the integration of the channel 2 signal was less than perfect, and that this might in part account for the calculations being a bit on the optimistic side. Since data on the behaviour of a 0.75 db echo had not been obtained, this surmise could unfortunately not be checked.

Again, the echo contributed by channel 4 was at a level of -1 db with respect to the primary sound, in those parts of the tests where all four channels were switched on. Conceivably therefore, a larger fraction of the energy of this channel was integrated than would have been the case if it had been at a level of 0 db ré the primary sound. This would make but small difference to the final value of the calculated answer, however, since the relative contribution of channel 4 was always insignificant.

The above surmise, if correct, leads to the unexpected result that two echoes, both of level 0 db ré the primary sound, and both occurring at a time shorter than 30 milliseconds after the primary sound, will contribute less to the overall intelligibility when they coincide than when there is some delay between them. For, if multiple echoes of 0 db level ré the primary tone are indeed integrated according to

the/....

the same law as a single echo, the two echoes will, when not coincident, be perfectly integrated (provided they both occur within the first 30 milliseconds). When coincident, however, they constitute in effect a single echo of level +3 db ré the primary signal, and in the light of the foregoing discussion, an echo of this level, coming at a time of say 20 milliseconds after the primary tone, will suffer less than perfect integration. It is regrettable that the apparatus available did not allow investigation of the accuracy of this interesting conjecture.

4.8. Comparison of results obtained with findings of other workers.

4.8.1. Results obtained by other workers:

As has been briefly mentioned in the introduction, various workers have already, by direct or indirect means, measured certain characteristics of the ear that depend on its integrating properties. By and large, these measurements were performed with pure tones or pulses rather than with speech as in the present case, but as the property being measured was the same in every case, a certain degree of correlation should be present.

The tests, methods and results of a few of these workers will now be considered in greater detail, and an attempt made to relate their results to those obtained by the present writer.

Békésy[16] has determined the increase in loudness of a pure tone, immediately after switching on. The experimental

technique/....

16) v. Békésy, G: "Zur Theorie des Hörens" - Physik. Zeitschr. November 1929, p. 743.

technique used by him was as follows:

A tone of 800 c.p.s. at a level of about 20 db above threshold * was presented to an observer through a pair of headphones. This tone was continuously interrupted by a motor-driven switch to give pulses of 800 cycle tone that had a duration of 100 milliseconds. By depressing a switch, the observer could switch over the same headphones to an 800 cycle test tone, the duration of which was preset by the experimenter. The observer then had to vary, by means of a calibrated attenuator, the intensity of this second tone until it appeared to him to be equal in loudness to the 100 milliseconds reference pulses.

The duration of the test pulses used by Békésy varied from 4.5 milliseconds to 45 milliseconds. Figure 28 shows his experimental results (solid line) redrawn with a linear time scale. The curve actually shows the relation $10 \lg L = k \sqrt{t}$, (L = intensity, k a constant and t = time elapsed since start of pulse) but Békésy's experimental results follow this law quite closely between 12 and 45 milliseconds. (See also figure 32).

To relate Békésy's experimental results to those of the present writer, reference must be made to figure 20. The derivation of this curve, which shows the integrated fraction of the energy contained in the echo, is given in section 4.2.2.

Now, if these results hold, not only for one single echo, but also for a great number of echoes of equal intensity - or, put in another way, if in the limit the integration of the ear is continuous and follows the law of figure 20 - then the integral of figure 20 would show the build-up of the loudness of a steady signal immediately after it has been switched on.

Figure 27/....

* Actually, Békésy gives the intensity as being "Hundertmal stärker als Hörschwelle" which is taken to refer to power, not pressure.

Figure 27 shows the integral of figure 20 in arbitrary units, as well as in db ré an arbitrary reference. This integral was obtained by measuring the area under the curve of figure 20.

The dotted line in figure 28 is the same as the curve of figure 27 and it will be seen that the agreement with Békésy's results is quite reasonable.

In his paper [16] Békésy states that the explosive character of pulses of short duration causes the loudness judgement to be high, since the observer is inclined to interpret the transition from "tone" to "explosion" as in itself being an increase in loudness. This might to some extent account for discrepancies below 15 milliseconds.

In his previous paper [2] Békésy found that the apparent loudness increased with time and reached a maximum at about 200 milliseconds. However, it is important to realise that this was for the case of a pure tone. In the case of speech, where the speech sounds have a duration of roughly 100 milliseconds, masking of speech sounds by adjacent ones probably occurs, and this might result in an integration period shorter than that for pure tones. Even for pure tones, the apparent loudness attained after 100 milliseconds is only $\frac{1}{2}$ db lower than the maximum at 200 milliseconds.

Garner [4] used repetitive pulses of which the duration and repetition rate could be varied independently. The pulses consisted of a keyed sine wave. Observers were required to indicate the hearing threshold for pulses of varying duration and repetition rate. Since Garner found that the frequency of the keyed wave made little difference to the shape of the curves obtained, he performed all but his preliminary tests with 1000 c.p.s. signal.

Using/....

Using random noise for masking purposes, Garner carried out his tests at a level of about 90 db re 0.0002 dynes/cm^2. In his paper, he gives the threshold of the pulses in terms of db signal-to-noise ratio, as a function of duration and/or repetition rate. These results could easily be converted to an equivalent level, since a lowering in threshold of, say 3 db, is equivalent to an increase in signal level of 3 db if the background noise is held constant.

Garner's results for a repetition rate of 10 pulses per second are shown in figure 29 (solid line). It shows the effective increase in level as the pulse length is varied. Since, with such a relatively slow repetition rate, the integration from pulse to pulse is very little, this curve in effect depicts the increase in loudness as a function of time, exactly as figure 28. Comparison with the integration curve (figure 27) obtained by the present writer, shows (triangles in figure 29) that the agreement is very good.

There is yet another way of looking at Garner's results. If the pulse length is kept constant, the apparent level, as far as the ear is concerned, depends on the repetition rate of the pulses. Now, if the results obtained with the word lists are in agreement with those found by Garner, the effective level of pulses of a given repetition rate and duration can be calculated with the aid of figures 20 and 27. To take a concrete example: Suppose the pulse length is 10 milliseconds and the repetition rate 50 per second. The pulses are then spaced 20 milliseconds apart. If the first pulse begins at t = 0, its contribution to the apparent sound level will be proportional to the integral, from t = 0 to t = 10 of the curve of figure 20. The contribution of the second pulse will be proportional to the integral from t = 20 to t = 30 of the curve of figure 20, and so on.

Since/...

Since figure 27 is the integral of figure 20, these contributions can be simply evaluated from the curve of figure 27, e.g. the contribution of the 4th pulse (80 to 90 milliseconds) is proportional to the value of the curve of figure 27 at 80 milliseconds, subtracted from its value at 90 milliseconds. Pulses occurring at a time t = 90 milliseconds contribute nothing at all.

It is also clear that the effective apparent level obtained is dependent not only on the spacing between the pulses, but also on their duration so that this method of summation takes into account both these variables.

Garner's results, redrawn to show effective level as a function of repetition rate, with pulse length as parameter, are shown plotted in solid lines in figure 30. The dotted line curves were calculated by the methods outlined above. It should be noted that not only is the increase in level as a function of repetition rate in good agreement, but also the level as a function of pulse length. The only notable exception to this latter agreement is the case of a tone of 1 millisecond duration. In this case the effective level measured by Garner is lower than indicated by the calculation.

A point worth noting is that for repetition rates lower than 10 tones per second (i.e. pulses spaced more than 100 milliseconds apart) Garner's results still show integration taking place. This integration is not very great, e.g. increasing the repetition rate from 5 per second to 10 per second only increases the effective level by about 1 db, which means that doubling the number of pulses per second has only added 20% to the effective level. However, this would seem to be in agreement with Békésy's results, and as explained in that connection this deviation from the results of the present writer, is probably attributable to the mutual masking of speech sounds on each other.

Munson 3)/....

Munson [3] also determined the rate at which the loudness of a pure tone increased with time, using very much the same technique as Békésy. However, in his case the reference tone had a duration of 1 second, and had to be adjusted by the observer until it sounded just as loud as the first, or test, tone the length of which was varied by the experimenter.

The dependence, as measured by Munson, of apparent loudness on duration for a 1000 cycle tone at a level of 20 db ré 0.0002 dynes/ cm^2 is given by the solid line curve in figure 31. The dotted line is again the curve obtained by integrating figure 20.

As can be seen, the agreement is good. However, the results obtained by Munson at a level of 80 db ré 0.0002 dynes/ cm^2 are in complete disagreement, as shown by the dashed line in figure 31.* It will be noted that this curve has an initial slope of 6 db per doubling in time, implying a pressure rather than an energy addition. The reason for this discrepancy is not clear.

4.8.2. Discussion of results of other workers in relation to those of the present worker:

As has been shown, the agreement obtained between the present work and that performed by the abovementioned three workers, is quite reasonable. The best agreement is with the results obtained by Garner.

It has also been shown that the results obtained by Békésy, Garner and Munson are in good mutual agreement, except those results obtained by Munson at higher levels. These latter results are in complete variance with those of the other abovementioned workers, and no convincing explanation for this difference can be advanced.

The/....

* Shown drawn with respect to an arbitrary reference level, chosen to facilitate comparison with other curves.

The method employed by Garner is probably the most satisfactory, since hearing threshold can be ascertained with higher consistency and accuracy than equal loudness. Also, as the test signals were of much lower intensity than the background noise, transient fatigue effects such as might be occasioned by loud interrupted pulses, are eliminated.

The greatest discrepancies between Békésy's results and those of the present writer occur at the shorter delay times, i.e. 10 milliseconds and less. It is suggested that Békésy's results could be in error here, for the following three reasons:

1. In his paper [16], Békésy states that at levels of 20 db and higher (referred to hearing threshold) the explosive character of short pulses causes the loudness judgement to be high, since the observer interprets the transition from "tone" to "explosion" as in itself being an increase in loudness, even though the amplitude be kept constant. This could partly account for the fact that Békésy's curve lies higher than that of the present writer, for short durations.

2. The curve obtained by Békésy is redrawn in figure 32. Although the presentation is different to his to facilitate comparison with the previous curves, the information is the same as in the version published by him. As will be seen, the spread of his experimentally obtained points increases as the durations become shorter. This suggests that the accuracy obtained by him at shorter durations is probably not very high.

3. Integration of the curve of figure 20 yields the curve of figure 27, which shows the build-up expected of a continuous tone. Now, conversely, differentiation of a curve showing the build-up of a continuous tone, should yield a curve that shows, as a function of time, which fraction of the incident energy is integrated by the ear. Békésy suggests that the relation

L = k/.....

$L = k\sqrt{t}$ describes the curve (figure 32), obtained by him. However, on differentiation, it is found that $dL/dt = \frac{1}{2}kt^{-1}/2$. This relation is plotted in figure 33 (solid line) and the curve of figure 20 is shown for comparison. An arbitrary value of 8 has been given to K in the above equation. As t becomes small, $\frac{dL}{dt}$ tends to infinity — this suggests strongly that the relation $L = K\sqrt{t}$ does not hold at small values of t.

The following important conclusions may be drawn from the foregoing results, comparisons and discussion:-

(1) The integration curve of figure 20 seems to be substantially correct.

(2) This integration curve seems to apply to pure tones of various durations, as well as to speech.

(3) The form of this integration characteristic is apparently not markedly affected by level in the range 20 - 45 db ré threshold.

(4) The integration curve seems to apply to single as well as to multiple echoes.

These conclusions are in disagreement with those drawn by Garner that for integration, the sound energy must be continuous, and must lie in an (unspecified) narrow band. From the present results, as well as their correlation with thoseof previous workers, it seems that the integration of a sound signal by the ear depends only on the time elapsed since its initiation and is fairly independent of its spectral composition and level.

The results obtained by the present writer are perhaps better than one would expect, considering the possible sources of error in subjective tests. It must be pointed out, however, that these tests were of the nature of threshold tests rather than subjective judgements, and therefore are inherently

more/....

more accurate. Furthermore, each point obtained is a mean value, representing the results of a large number of observations. This would tend to minimize random errors, but it is of course not impossible that systematic errors might have occurred. In view of the precautions taken, however, it is hoped that this is not the case.

5. POSSIBLE APPLICATIONS TO ARCHITECTURAL ACOUSTICS

5.1. General review:

While there are many desirable and undesirable acoustical properties pertaining to enclosed spaces such as auditoriums, broadcast studios, cinemas and the like - some of these properties being desirable for one type of enclosure and possibly detrimental in another - it has hitherto not proved possible to define all these properties physically, or even to measure them objectively. Practically the only physical measurement which has proven to be of sufficient repeatability to warrant its inclusion as a factor in the assessment of the "goodness" of a hall, is the measurement of reverberation time.[17, 18)]

Although there is an optimum reverberation time for a hall of given volume [19)] (the optimum being a compromise between different factors, which, often conflicting, depend also on whether speech or music or both will be heard in the hall) the correlation between the reverberation time and merit of a hall leaves much to be desired from the designer's point of view. Measurements in a number of halls [18)], for example, show that while halls that are generally considered good have reverberation times close to the optimum value specified by Knudsen, the inverse does not, unfortunately, hold good. A correct reverberation time may therefore be considered as a necessary, but not sufficient condition for satisfactory acoustics.

The work of Haas, quoted in Section 1, has made it possible to employ another approach toward the solution of the

problems/....

17) Kuhl, W.: "Nachhallzeit grosser Musikstudios" - Akustische Beihefte (2), (1954), p. 619.

18) Parkin, Scholes and Derbyshire: "The Reverberation times of ten British Concert Halls" - Acustica Vol. 2, No. 3, pp. 97 - 100 (1952)

19) Knudsen, V and Harris, C: "Acoustical designing in architecture" p. 194 (Wiley)

problems connected with the acoustical design of halls. While it had been generally realised that the ratio of direct to reverberant sound and the arrival time and intensity of echoes profoundly influenced the intelligibility of speech and the character of musical sound and also determined the "definition" "balance", "blend", "fullness of tone" (Parkin [20] mentions about six terms of this nature which, while subjectively clear are very difficult to define physically) design intended to attain these objectives frequently had to depend on intuition and previous experience, and final assessment of whether the design objective had been realised depended on opinions which, even though voiced by experienced musicians and architects, could be widely divergent [20].

Bolt and Doak [11] have suggested a criterion for the transient acoustic response of rooms, based on the results of Haas and previous work done by themselves. This criterion relates the intensity at any instant of the decaying sound in a room to the time elapsed after the cessation of the original sound. Muncey, Nickson and Dubout [12] investigated the effect of reverberation time on the disturbance caused by a single artificial echo, and found that for rooms with short reverberation times a more stringent criterion than the one proposed by Bolt and Doak was required. Measurements performed by Parkin and Scholes [21] in a theatre seem to indicate that the Bolt and Doak criterion is somewhat too stringent.

The measurements carried out by the present writer in an anechoic chamber indicate that equal loudness of primary tone and echo is attained at lower levels of echo, for delay times exceeding 20 milliseconds, than found by Haas. It can therefore be argued that disturbance due to the presence of

the/...

20) Parkin, P.A. and others: "Acoustics of the Royal Festival Hall" - Acustica (3), pp. 9 and 10.
21) Parkin and Scholes: "Recent Developments in sound reinforcement systems" - Wireless World (57) pp. 44 - 50

the echo will occur at lower echo levels than found by him, and this is in agreement with the findings of Muncey, Nickson and Dubout. Clearly, therefore, the Bolt and Doak criterion is affected by reverberation time, and becomes more stringent as the reverberation time becomes shorter.

Practical use has been made of the so-called Haas effect (masking of an echo by the primary tone) in the reinforcement of sound [21]). In general, sound reinforcement is required for speech only, and finds its main application in very large or poorly designed auditoriums, where the normal energy output of the human voice is inadequate, although the speaker is still audible. Public address systems could also profitably employ this system, which in principle consists of delaying the amplified signal to ensure its arrival at the listener after the arrival of the direct sound. In the case of public address systems, it could be used in conjunction with directional loudspeakers to obviate disturbing echoes.

5.2. Application to sound reinforcement systems:

The principle underlying sound reinforcement with a delayed signal is that, provided the delay is correctly chosen, naturalness is preserved even with the amplified signal at a level (at the position occupied by the listener) of about +8 db with respect to the direct sound. A considerable increase in intelligibility can therefore be obtained, while still preserving the illusion that all the sound comes from the speaker. Moreover, the direction from which the amplified sound reaches the listener is not important, so that a suitable location for the loudspeaker can usually be found. The integration by the ear of a single echo of level 0 db and of +5 db with respect to the primary sound, as determined by the present writer and shown in figures 19 and 22 is useful in this connection, as it gives an indication of how the total

effective/...

effective level depends on the delay time. Especially in the case of a +5 db echo (and presumably even more so for the case of echoes at a higher level) it would seem desirable in the interest of maximum intelligibility to keep the delay as short as possible without sacrifice of naturalness.

One of the advantages of delayed sound reinforcement, quite apart from the fact that it has a much more natural quality than the more orthodox sound amplification systems without delay, is the fact that higher intelligibility of speech can be attained than with the undelayed systems.

If we consider the case of a listener seated at a distance of say, 55 feet from the speaker, close to a loudspeaker of the amplifying system, the following conditions exist:

If the microphone is close to the speaker, undelayed sound amplification will result in sound from the loudspeaker reaching the observer at a time about 50 milliseconds ahead of the natural sound. Neglecting the effect of room reverberation, the signal heard by the observer will consist of a primary sound coming from the loudspeaker, and a single echo coming from the (human) speaker. Taking the sound level (as measured at the position of the observer) due to the human speaker as reference (0 db) the effective levels due to integration by the ear can be found from figures 19, 21 and 22 respectively for the cases where the loudspeaker is adjusted to levels (re the natural sound) of 0, -5 and +5 db. Reference to the figures mentioned shows the effective levels to be as follows (all with respect to the level of the natural sound):

Loudspeaker level db.	Effective level db.
- 5	- 1
0	+ 1.7
+ 5	+ 5.8

The/....

The first case corresponds to a primary sound with +5 db echo, the second to primary sound plus 0 db echo, and the third to primary sound with -5 db echo. Note that in the first case, the amplification system actually reduces the effective level. As will be shown later, this effect becomes even more pronounced when reverberation is taken into account.

If, on the other hand, a delay were introduced in the amplification system, and adjusted to a value which would ensure that the sound from the loudspeaker reached the observer at a time of about 10 milliseconds after the sound from the human speaker, radically different conditions would be obtained. The sound from the loudspeaker would now constitute the echo, and the effective levels, again taking as reference the sound from the human speaker, would be:

Loudspeaker level db	Effective level db
- 5	+ 1.2
0	+ 3
+ 5	+ 5.6

An effect which has not been considered in the above example is that in the undelayed case, the sound from the loudspeaker will appear to the observer to be an unwanted signal rather than a primary sound, since he will be concentrating, aurally as well as visually, on the human speaker. It is likely, therefore, that due to the subjective discrimination possible with binaural hearing, the undelayed amplification system will be less effective than appears from the calculation. In the case of delayed sound reproduction, this consideration does not apply, since in a correctly adjusted system all the sound appears to come from the human speaker.

It will be noted that for amplified sound of higher level than the natural sound, the delayed system appears to

be less efficient than the undelayed system. However, the above discussion pertains to conditions existing in the absence of natural echoes. When the effect of reverberation is considered, the great advantage of delayed sound reinforcement becomes clear:

In the absence of any sound reinforcement whatsoever, the primary sound reaching the observer consists of the direct sound from the source. This is followed by a number of echoes decaying (in the ideal case) exponentially, and constituting the reverberant sound. Now, the useful sound energy is contained in the signal arriving at the observer during the first 90 milliseconds. The direct sound and the first 90 milliseconds of reverberant sound therefore constitute the useful signal, and the rest of the reverberant sound acts as noise.

With conventional loudspeaker systems, the primary sound reaching the observer emanates from the loudspeaker. This is followed after (for argument's sake) 30 milliseconds by the direct sound from the human speaker which in turn is followed by the echoes due to the reverberant sound. It is assumed that the loudspeakers are highly directional or close to the listener, and do not add materially to the reverberant sound. Again, the useful signal is contained in the first 90 milliseconds. In this case, therefore, the useful signal is made up of the amplified sound, plus the direct sound from the speaker, plus only the first 60 milliseconds of the reverberant sound - since the ear now begins to integrate 30 milliseconds earlier than in the case where natural sound only was present. The remaining reverberant sound adds nothing to the intelligibility and acts as noise. In effect, therefore, the useful contribution of 30 milliseconds of the reverberant sound has been lost, and the reverberant noise has been increased.

If, finally, the case is considered where the amplified signal has been delayed to occur say, 20 milliseconds after the direct sound from the speaker has reached the observer, the position is as follows: The direct sound from the human speaker, plus the signal from the amplifying system, plus the first 90 milliseconds of reverberant sound all contribute toward the useful signal, and the reverberant sound occurring after the first 90 milliseconds acts as noise.

From the above discussion it is clear that for equal loudspeaker output, the delayed system will yield higher intelligibility than the undelayed system, because a greater fraction of the reverberant sound acts as useful signal, and a smaller fraction of the reverberant sound acts as noise. This surmise appears to be borne out by results obtained in practice.

(1) Parkin [22] reports consistently higher articulation percentages with delayed speech reinforcement than with the same system without delay. This is in agreement with the above explanation.

(2) Results obtained in the House of Assembly in Cape Town [23] are shown in figure 34. In these tests, the observers were seated at a distance of about 30 feet from the speaker. Small loudspeakers were affixed to the seats behind the observers, and delayed or undelayed signal could be fed to these loudspeakers. The artificial delay introduced was about 50 milliseconds. The two curves show clearly the improvement in articulation obtained with the delayed system. Note also the initial decrease in intelligibility that occurs with undelayed speech reinforcement at low loudspeaker levels.

5.2. Application/....

22) Parkin, P.H. and Taylor, J.H.: "Speech reinforcement in St. Paul's Cathedral" - Wireless World, Feb. 1952, pp. 54 - 57 and Wireless World, March 1952, pp. 109 - 111.

23) "Akoestiek van die Suid-Afrikaanse Volksraadsaal", C.S.I.R. Report A57/48194/1 of June 1954.

5.3. Application of experimental results to the computation of articulation percentages:

The method of articulation tests is one which yields excellent results in the testing of halls, especially those intended principally for speech. Observers, placed at different positions in the hall, are required to write down test lists which are read to them. Provided enough observers are available, very useful information can be derived from these tests. The tests may also be used to determine the effect of acoustical treatment, simply by comparing results obtained before and after the alterations.

A few different types of articulation lists exist - monosyllabic test lists, spondees, sentence tests and the like. The greatest handicap shared alike by all of them is the necessity of employing a large number of observers if consistent results are to be obtained. For reliability, tests carried out in halls, where repeatable results are required at quite a few different positions, require anything from fifty observers and upwards. In many cases, suitable observers are not available, or not available in sufficiently large numbers.

French and Steinberg [24] have described a method by which it is possible to calculate articulation percentages from physical measurements. In its essentials, the method consists of level measurements over twenty critical bands each of which will, under optimum conditions, contribute an equal amount to the overall intelligibility. By measuring the useful signal and the noise in each band, it is possible to compute the percentage articulation that would be obtained with an experienced observer.

The/....

24) French, N.R. and Steinberg, J.C.: "Factors governing the intelligibility of speech sounds", Journ.Acous. Soc.Amer. Vol. 19, pp. 90 - 119.

The beauty of the method lies therein that physical measurements suffice to yield an answer in terms of what is really a subjective measure, and that the consistency of the answers depends only on the physical measurements and not on the idiosyncracies of individual observers. Unfortunately, application of the method to building acoustics poses many problems, in particular as a result of echoes which are invariably present to a greater or a lesser degree. It could be possible, however, that the method might be successfully employed if use were made of the integration characteristics of the ear as found by the present writer. The total effective level in each critical band would then be the resultant of the direct, or primary, sound and the following echoes. The level, as a function of time, of the primary sound and echoes will have to be measured at the spot for which the percentage articulation is to be calculated. Integration of primary sound and echoes can then be performed, in accorance with the results described in a previous section, to obtain the total effective signal in each band. Sound occurring later than 90 milliseconds after the primary tone can be regarded as noise, and added, on a power basis, to any other noise that may be present in that particular band. A considerable amount of work still has to be done, especially as far as the instrumentation side of the problem is concerned, before this method can be applied to architectural acoustics; but if it is successful, it should prove quite a big step forward in the assessment of the acoustics of halls.

APPENDIX

CALCULATION OF OPTIMUM REVERBERATION TIME FOR ROOMS IN WHICH INTELLIGIBILITY OF SPEECH IS OF PRINCIPAL INTEREST

Although reverberation time is not the fundamental characteristic of a room that determines the intelligibility of speech and the quality of music, it is a secondary effect which can be measured relatively easily and which has for a long time been taken as a measure of the suitability of rooms. A further merit of this characteristic of a room is the fact that it can fairly easily and relatively accurately be controlled by the designer. With these merits and in spite of all the latest work in the field of architectural acoustics, reverberation time will most probably be the primary factor in the design of rooms for some time to come.

As far as music is concerned, the reverberation time affects the quality of the musical tone but the subjective results - blend, presence, definition and the like - are apparently so highly divergent from person to person that correlation with physical measurements is extremely difficult. The optimum reverberation times for rooms of different volumes are therefore based mainly on judgements of whether a given hall with a given reverberation time sounds good or bad. Parkin and others [18] found that the optimum times given by Knudsen and Harris [19] fall very close to those measured by them in halls usually designated as "good". Recent work by Kuhl [17] suggests that the type of music rather than the hall volume, determines the optimum reverberation time.

Concerning rooms designed principally for speech, the problem is much more clear-cut. The reverberation time should be of such a nature that naturalness of speech is preserved (e.g. undue discrimination against any frequencies in the speech frequency spectrum should be avoided) and its value chosen to yield maximum intelligibility of speech.

If/....

If the two extremes are considered - a hall with no reverberation at all on the one hand, and a hall with a very long reverberation on the other - it will be seen that both cases will yield low intelligibility if the halls are large. In the case of the hall with no reverberation, free field conditions exist as far as speech propagation is concerned, with the result that the sound pressure level at positions far from the speaker will be too low for satisfactory understanding of speech. On the other hand, if the reverberation time is too long, intelligibility will be impaired by delayed echoes which act as masking noise. Apparently, therefore, a reverberation time having a value somewhere between these two extremes can be found that will yield a maximum useful signal to noise ratio and which can therefore be considered as the optimum value for that given volume.

In the calculations which follow of optimum reverberation time as a function of volume, certain assumptions were made, which will now be given.

(1) The dimensions of the rooms are: length 1.5D; width 1.0D; height 0.5D. Consequently the volume is 0.75 D^3 and the total surface area is 5.5 D^2.

(2) The distance from observer to speaker is 1.0 D.

(3) The diffusion is perfect, and the sound decay is a continuous exponential function, so that the sound pressure p, at any given time t after cessation of signal is given by $p = p_o E^{-\frac{13.8t}{T_{60}}}$ where T_{60} is the reverberation time in seconds, and p_o the steady state sound pressure reached before cessation of the signal.

(4) The reverberation time is independent of frequency.

(5) The masking properties of reverberant sound caused by speech are the same as those of the random noise used in an experiment which will be described later.

(6) The/....

(6) The speaker has an acoustical power of 50 microwatts and is a perfect point source (i.e. omnidirectional).

(7) The noise generated by the audience is uniformly distributed over the entire floor area, and has a density of 3.16×10^{-12} watts/cm^2.

(8) The steady-state level is reached very rapidly, so that initial build-up of the level can be neglected.

(9) The integration characteristic of the ear as depicted in figure 20 holds for multiple echoes also, and is independent of the ratio echo level to primary sound level, as well as independent of absolute level.

Calculations of the contribution of the reverberant sound to the signal and noise components of the sound reaching the observer were now made as follows:

The intensity of the reverberant sound is given by $W = WoE^{-13.8t/T_{60}}$ watts per cm^2. Over the first 30 milliseconds, the useful energy is given by $E_0^{30} = Wo\int_0^{0.03} t^{-13.8t/T_{60}}\,dt$. Over the period t = 30 milliseconds to t = 90 milliseconds, the useful energy is given by $E_{30}^{90} = Wo\int_{0.03}^{0.09} K(t)E^{\frac{-13.8t}{T_{60}}}\,dt$ where K(t) is a function of time given by the curve of figure 20. This part of the integration was performed graphically. The rest of the reverberant sound occurring after 90 milliseconds was regarded as noise, the energy being given by the relation $E_{NR} = Wo\int_{0.09}^{\infty} E^{\frac{-13.8t}{T_{60}}}\,dt$. The total energy in the reverberant sound is given by $E_{TR} = Wo\int_0^{\infty} E^{-13.8t/T_{60}}\,dt$.

The sound intensity at the position occupied by an observer D feet from the speaker, disregarding the noise due to the audience, is $Wo = \frac{4Ws}{R}$ watts/cm^2 25) due to the reverberant sound (Ws is the power output of the speaker in watts, $R = \frac{S\alpha}{1-\alpha}$ where S is the total surface area of the room in square centimeters, and α the absorption coefficient).

The/...,

25) Beranek, L: Acoustic measurements (Wiley) p. 673.

The parameter R can be calculated from T_{60} using the Eyring formula $T_{60} = \frac{0.05\ V}{-S\ \ln\ (1-\alpha)}$.

It is next assumed that as far as the reverberant sound is concerned, the ratio useful power to total power is equal to the ratio useful energy to total energy, i.e. $Wu/Wo = E_{UR}/E_{TR}$ where $E_{UR} = E_{0}^{30} + E_{30}^{90}$. The useful reverberant power is therefore found from the previous calculations: $WuR = E_{UR}/E_{TR} \times Wo$. To find the total useful power, the contribution from the direct sound must be added. This is $W_D = \frac{Ws}{4\pi r^2}$ [25] where r = 12 x 2.54 D centimeters. The total useful power is therefore $W_T = WuR + W_D$. Similarly, the reverberant noise power is $W_{NR} = E_{NR}/E_{TR} \times Wo$ watts/cm^2 to which must be added the noise due to the audience, which also consists of two components: The direct component, which is 3.16×10^{-12} watts/cm^2 and the reverberant component given by $W_{AR} = \frac{3.16 \times 10^{-12} \times 1.5 D^2 \times 4 \times 144 \times 2.54^2}{R}$ watts/cm^2. The total noise is hence $W_{NT} = W_{NR} + 3.16 \times 10^{-12} + W_{AR}$ watts/cm^2.

One more factor must still be known to enable the determination of optimum reverberation time: the signal level required with a given noise level, to obtain a certain arbitrary percentage articulation. To evaluate this, an observer in the anechoic chamber was presented with word lists in the presence of random noise. The signal level required to attain a 50% articulation score at different levels of masking noise was found, and the results are shown in figure 35. All levels are given in db referred to hearing threshold, which is assumed to be 10^{-15} watts/cm^2 for the purpose of these calculations.

Now, knowing the total noise in the hall, the curve of figure 35 could be used to determine the signal level required for 50% articulation. By assigning various values to T_{60} and calculating the total useful signal and the total

noise/....

noise for each value, the total signal and the total noise can be plotted as a function of T_{60}. By plotting on the same axes the signal level required for 50% articulation at each value of T_{60}, the excess of useful signal over that required for 50% articulation can be obtained. The value of T_{60} where this excess is a maximum, is then the optimum reverberation time.

An example of the calculations carried out may clarify matters. The speaker is assumed to radiate 50 μW under all conditions. When D = 51 feet, V = 100,000 cubic feet and S = 14,300 square feet.

When $T_{60} = \frac{1}{2}$ second

Wo (reverberant sound due to speaker) = $\frac{4 \times 50 \times 10^{-6}}{R}$ watts/cm^2

$T_{60} = \frac{0.05V}{-S \ln(1-\alpha)}$ i.e. $-\ln(1-\alpha) = \frac{0.05 \times 10^5}{0.5 \times 14,300}$

from which $\alpha = 0.504$

Thus $R = \frac{S\alpha}{1-\alpha}$ square feet, or 13.5×10^6 square cms.

Hence $Wo = \frac{200 \times 10^{-6}}{13.5 \times 10^6}$ watts/cm^2

The useful energy in the first 30 milliseconds is given by $E_0^{30} = W_o \int_0^{0.03} t^{\frac{-13.8t}{0.5}} dt = 0.0204\ W_o$ joules/cm^2.

Useful energy in the interval 30 to 90 milliseconds is given by $E_{30}^{90} = Wo \int_{0.03}^{0.09} K(t) t^{\frac{-13.8t}{0.5}} dt = 0.0064$ Wo joules/cm^2.

The <u>total energy</u> in the reverberant sound due to the speaker is $E_{TR} = Wo \int_0^{\infty} t^{\frac{-13.8t}{0.5}} dt = 0.0362$ Wo joules/cm^2.

The <u>useful reverberant energy</u> is (0.0204 + 0.0064) Wo joules/cm^2. The level is therefore $\frac{0.0268}{0.0362}$ Wo watts/cm^2,

i.e. $WuR = 10.95 \times 10^{-12}$ watts/cm^2. To this must be added the <u>direct sound</u> from the speaker.

$$W_D = \frac{50 \times 10^{-6}}{4\pi \times 51^2 \times 12^2 \times 2.54^2} = 1.65 \times 10^{-12} \text{ watts/cm}^2$$

$\therefore$ <u>Total useful signal</u> = 12.6×10^{-12} watts/cm^2.

Reverberant <u>noise energy</u> due to the speaker is given by

$W_{NR} = Wo \int_{0.09}^{\infty} E^{\frac{-13.8t}{0.5}} dt = 0.00302$ Wo joules/cm^2

$\therefore$ Noise/....

$\therefore$ Noise level due to speaker $= \frac{0.00302}{0.0362} W_o = 1.24 \times 10^{-12}$ watts/cm^2

Reverberant noise due to the audience is given by

$$W_{AR} = \frac{3.16 \times 10^{-12} \times 1.5D^2 \times 144 \times 2.54^2 \times 4}{13.5 \times 10^6} = 3.38 \times 10^{-12} \text{ watts/cm}^2$$

Direct noise due to the audience is 3.16×10^{-12} watts/cm^2

$\therefore$ total noise $= (1.24 + 3.38 + 3.16)10^{-12}$ watts/cm^2 $= 7.78 \times 10^{-12}$ watts/cm^2

Hence, referred to 10^{-16} watts/cm^2, the useful signal level is 51 db and the noise level 48.9 db.

These calculations were carried out for different values of reverberation time. In the case of $T_{60} = 0$ it was assumed that only half the noise power generated by the audience reached the observer, the other half being absorbed by the floor. The signal and noise levels, in db re 10^{-16} watts/cm^2 were then plotted as a function of reverberation time and the signal level required at each noise level to produce 50% articulation was found from figure 35, and plotted on the same axes. The difference, in db, between the actual signal level and the level required for 50% articulation was also plotted, and the value of T_{60} where this attained a maximum was taken as the optimum reverberation time.

The values obtained for a hall of 100,000 cubic feet are shown in figure 36. The optimum reverberation time is seen to be about 0.6 seconds. Repeating the calculations described above for different assumed values of volume, enabled the curve of optimum reverberation time as a function of volume to be drawn, figure 37. The optimum curve due to Knudsen [26] is also shown. It is of interest to note that Knudsen's curve was calculated from the following data, all obtained experimentally:

(1) The/....

26) Knudsen, V.O.: Architectural Acoustics (Wiley) p. 394.

(1) The decrease in percentage articulation with increase in reverberation time. This relation was obtained from articulation tests performed in halls of approximately the same shape and volume but different reverberation times. Correction was made in each case to give the percentage articulation which would have been obtained with an average sound level of 70 db. From this corrected curve, the reduction factor for the articulation percentage for increasing reverberation time was computed; this factor is designated by Kr.

(2) In a large hall, a speaker tends to speak louder than in a smaller hall. Measurements were made of the acoustic output of various speakers in halls of different sizes, to obtain the relation between average speech power and hall volume.

Now, assuming a given volume, the average acoustical power of the speaker was determined from (2), and the resultant level for different values of reverberation time could be calculated. This level was expressed as a fraction K_L of a level of 10^{-9} watts/cm^2 (which is regarded as yielding maximum articulation under optimum conditions).

The product Kr K_L is then indicative of the articulation to be expected, and the value of reverberation time for which this product was a maximum was taken as the optimum for the particular volume for which the calculations were made.

Knudsen's curve therefore takes no direct account of the ambient noise due to the audience and of the direct sound which, in the case of small halls with a short reverberation time, may be an appreciable fraction of the total useful energy received by the listener.

As will be seen from figure 36 the total noise present is an important factor in the determination of the optimum reverberation time. The value of the noise due to the audience would appear to have been chosen rather high; but

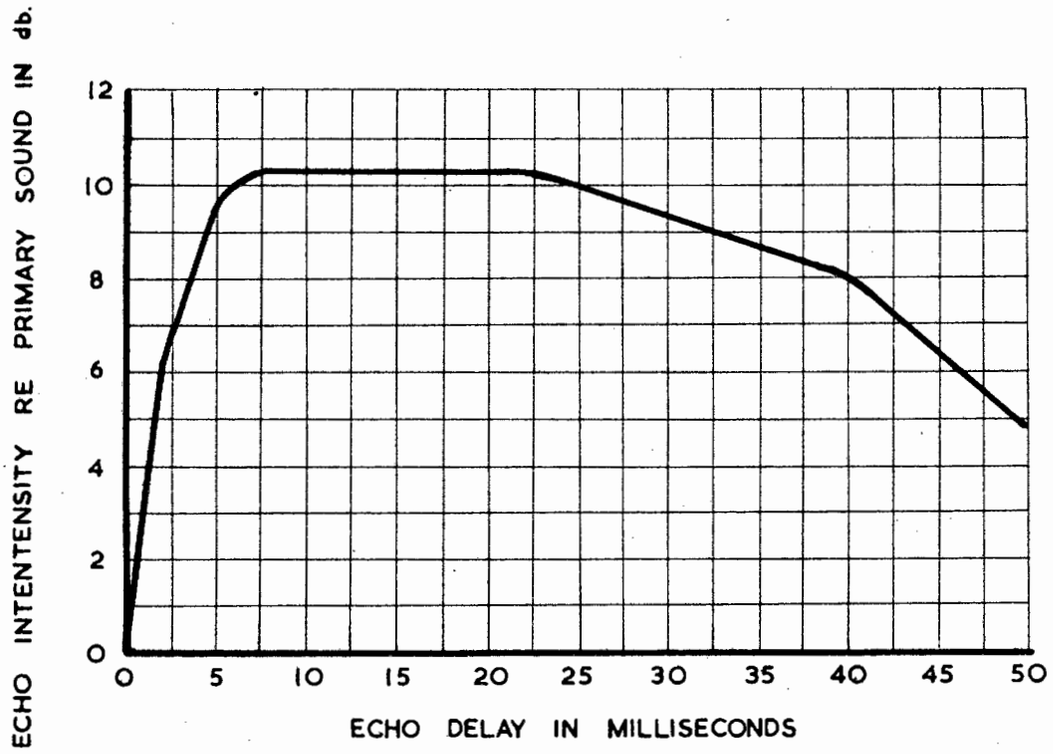
at/....

at the time when the calculations were carried out, this value was chosen purely for convenience and the implications only came to be realised after most of the rather laborious arithmetic had been done.

Although it does not appear as if a slightly lower noise level would affect the optimum reverberation times calculated very much, it is proposed to glean more accurate information as to its value from measurements in halls with an audience present. The optimum reverberation times can then be recalculated in the light of these results.

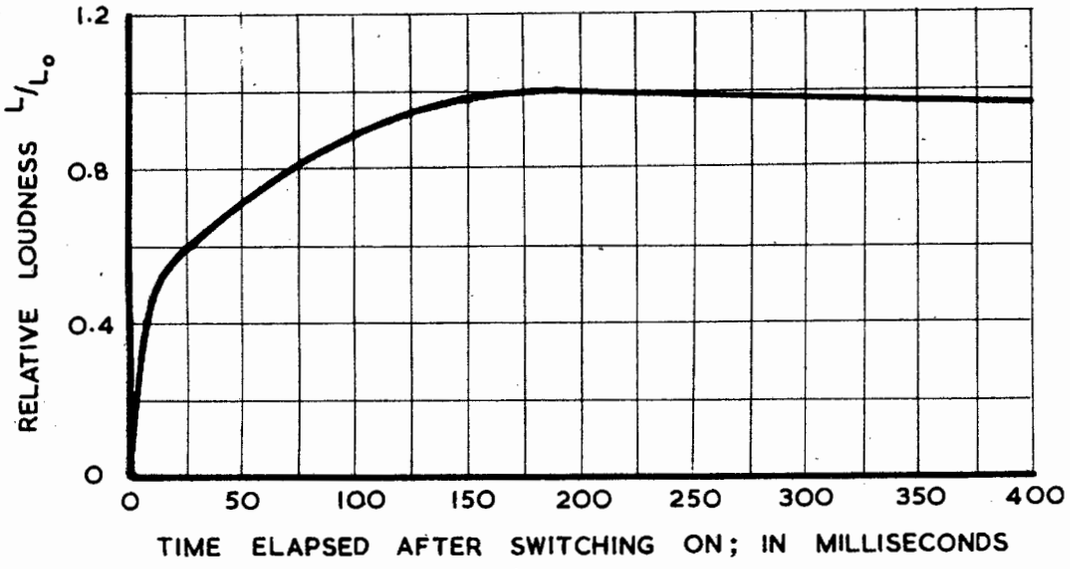
Another point to be noted is that the optimum reverberation time as calculated is not very sharply defined. This is also the case in the calculations given by Knudsen, and is in accordance with the results obtained in practice that the optimum reverberation time of a room is not very critical.

As has been stated earlier on, perfect diffusion and uniform decay were assumed in calculating the optimum reverberation times. In a well designed auditorium, however, reflective and absorptive surfaces are so arranged as to ensure that the first reflections reach the audience within the integration period and delayed echoes are minimized, with the result that the above assumptions no longer hold. Under such conditions it is possible that maximum speech intelligibility will require a reverberation time more in agreement with Knudsen's results - if one can still talk of a definite reverberation time of a hall under such circumstances.



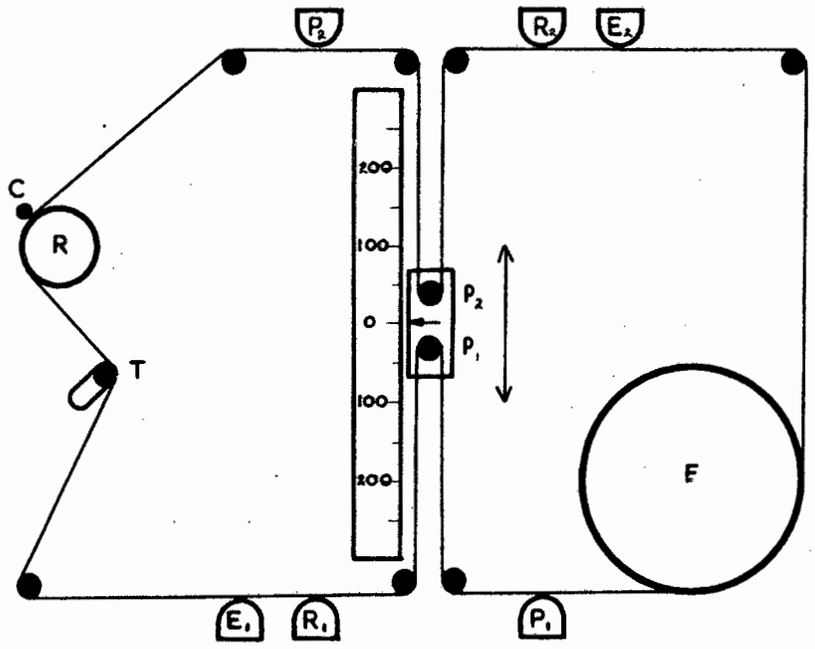
Curve showing the relative intensities for equal loudness as a function of the echo delay time. (after Haas)

FIG. 1



Increase in loudness of an 800 cycle tone after switching on. (after von Bekesy)

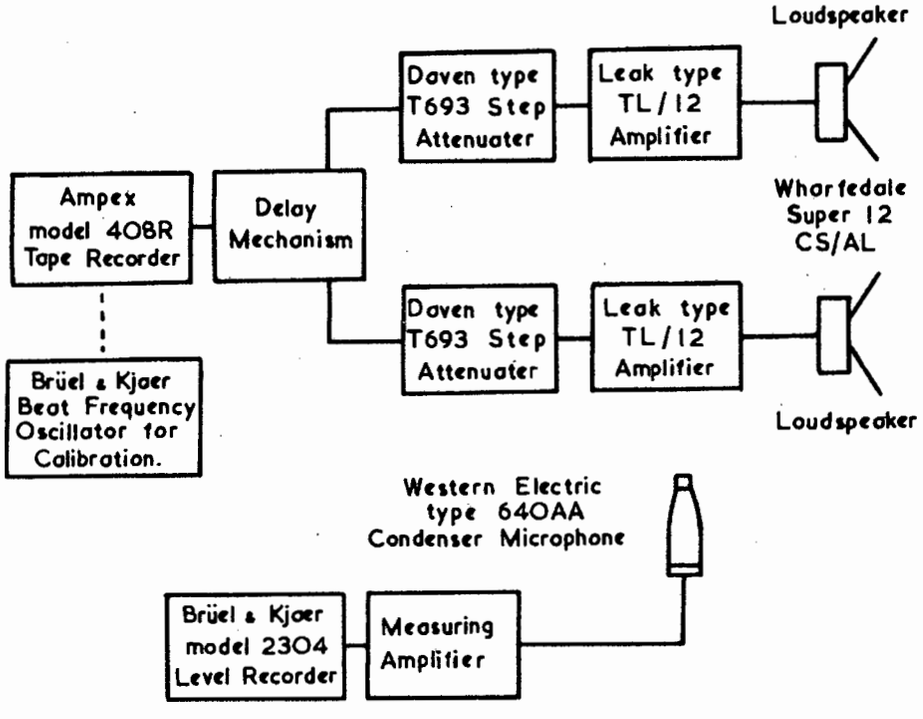
FIG. 2



E, E_2: Erase heads. C: Driving capstan.
R, R_2: Record heads. R: Pressure roller.
P, P_2: Playback heads. F: Flywheel.
p_1 p_2: Pulleys on movable carriage.
T: Tensioning pulley.

Diagram of Delay Mechanism.

FIG. 3



Block diagram of equipment used in tests with a single delayed echo.

FIG. 4

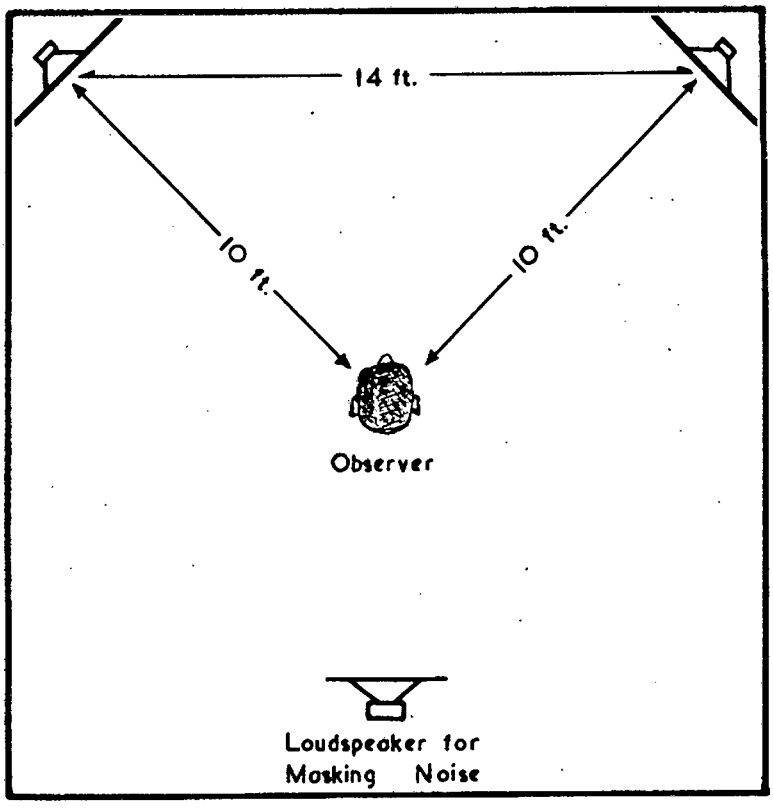
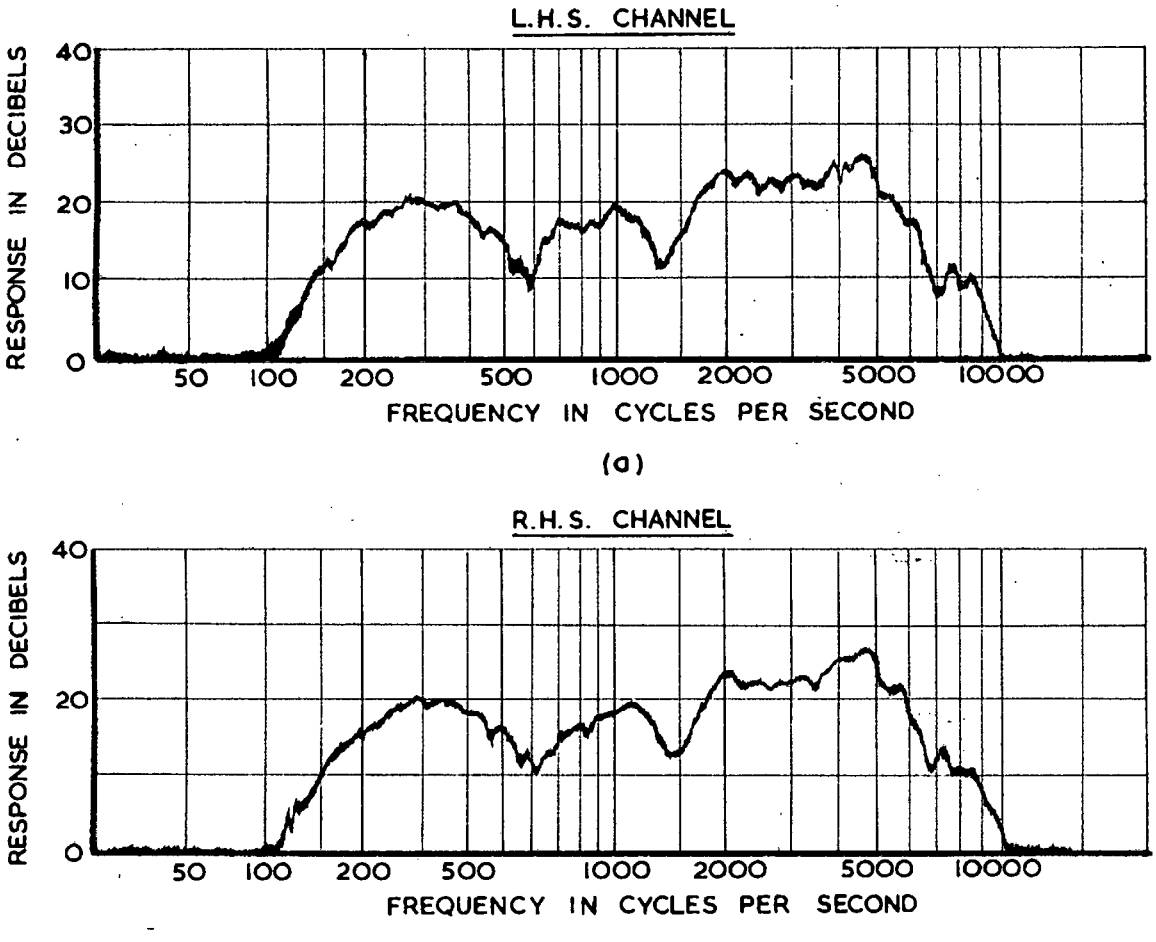


Diagram showing relative positions of observer and loudspeakers in anechoic chamber, for tests with a single echo.

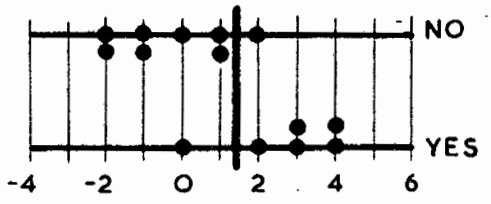
FIG. 5



Overall frequency response of two channels used in tests with a single delayed echo.

FIG. 6

Level of primary sound: 50 db ré hearing threshold.
Delay time: 20 milliseconds.
Observer: D. M.

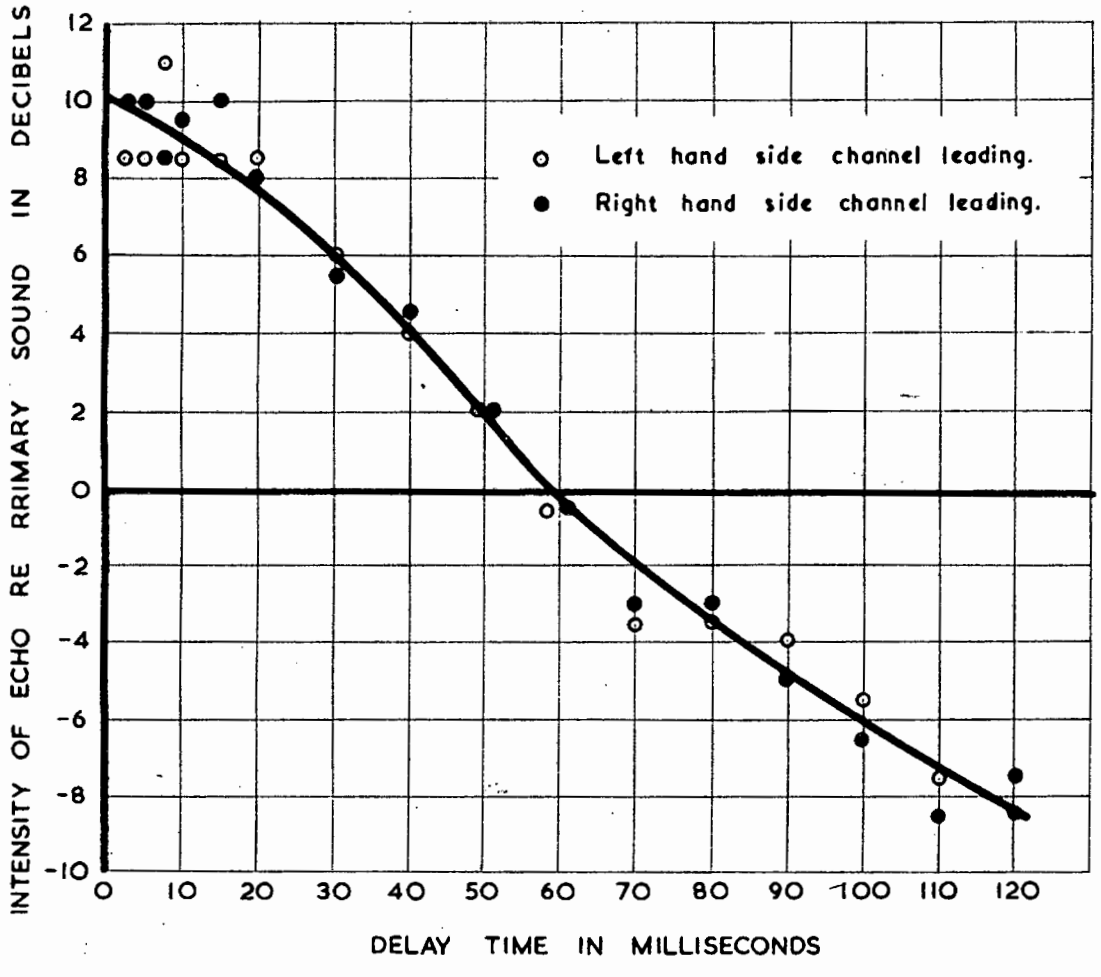


LEVEL OF ECHO RE PRIMARY SOUND
(DECIBELS)

Threshold of perception judgements: notational method.

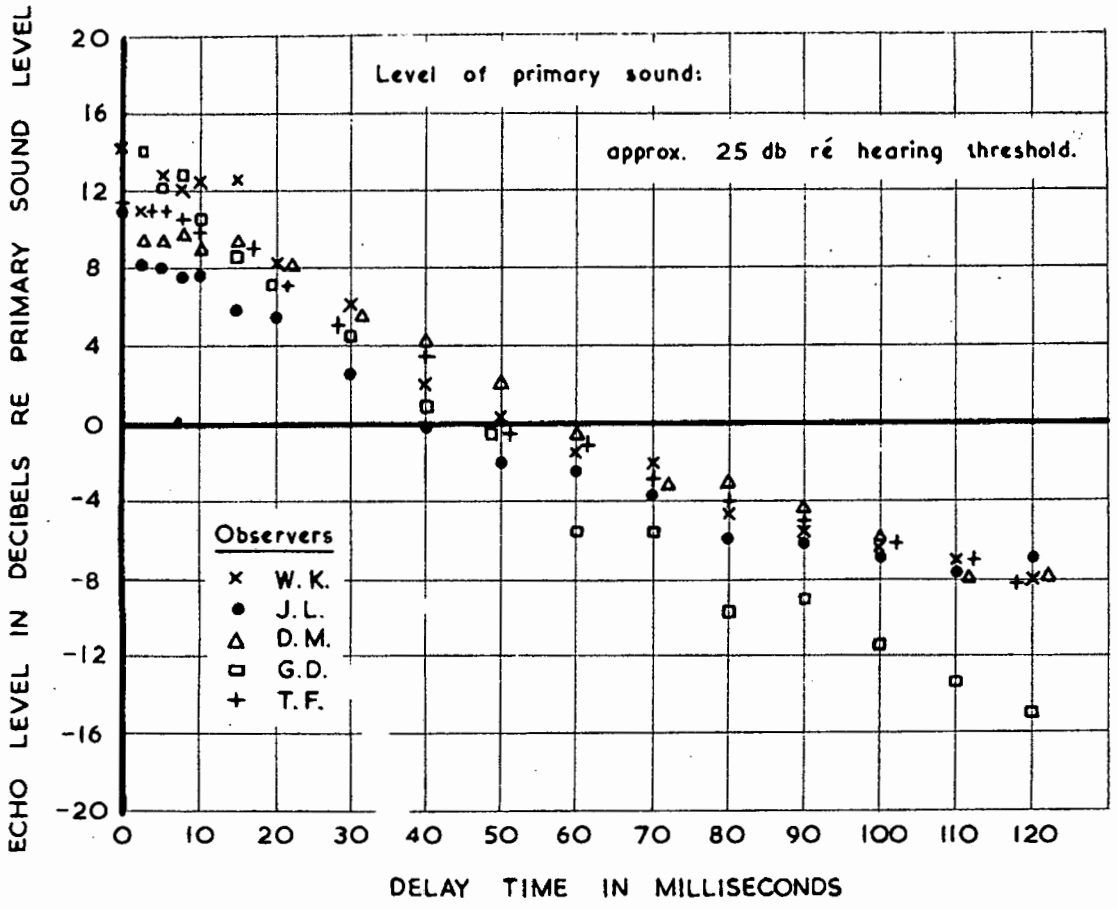
FIG. 7

Level of primary sound: 25 db ré hearing threshold.
Observer: D. M.



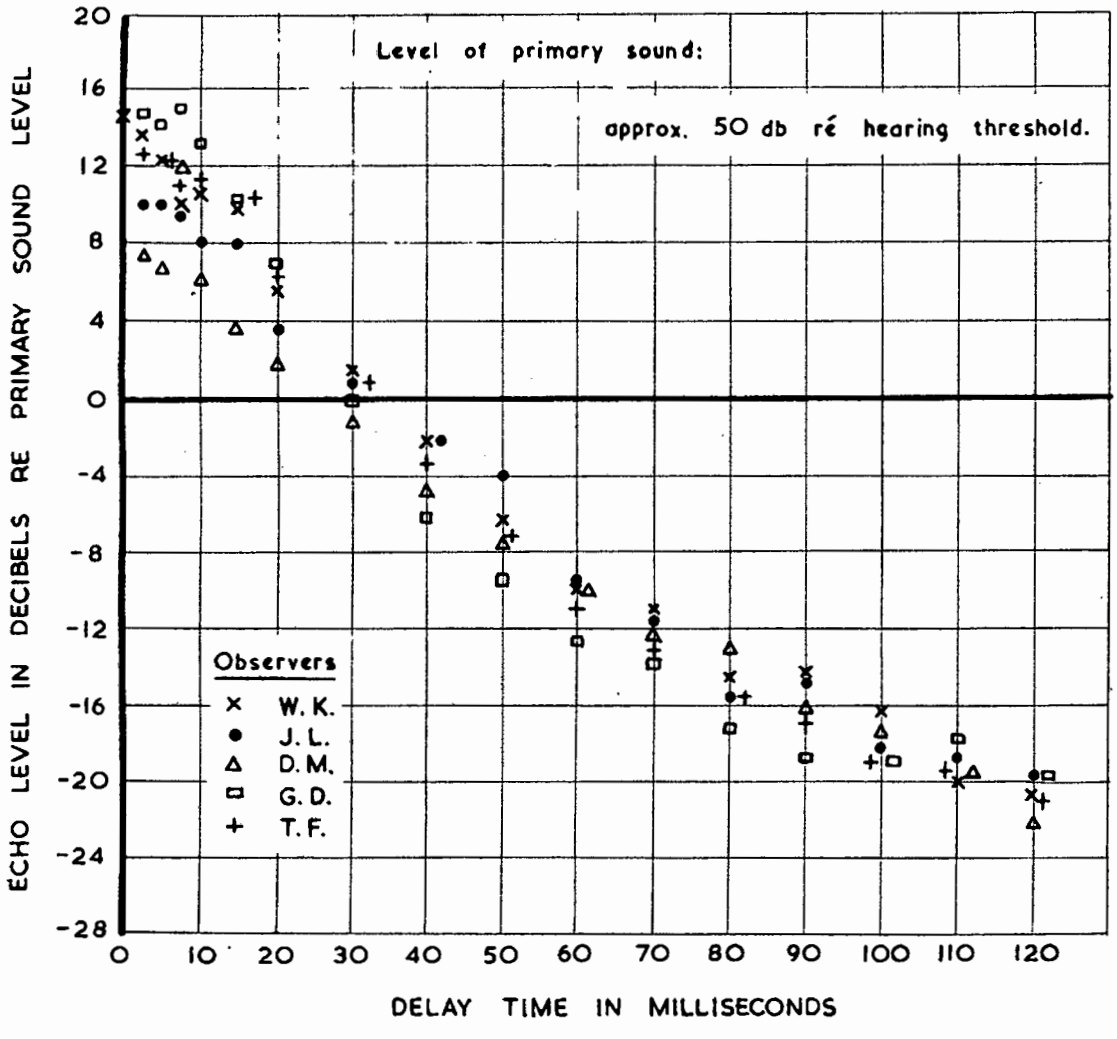
Threshold of perception contour obtained with one observer.

FIG. 8



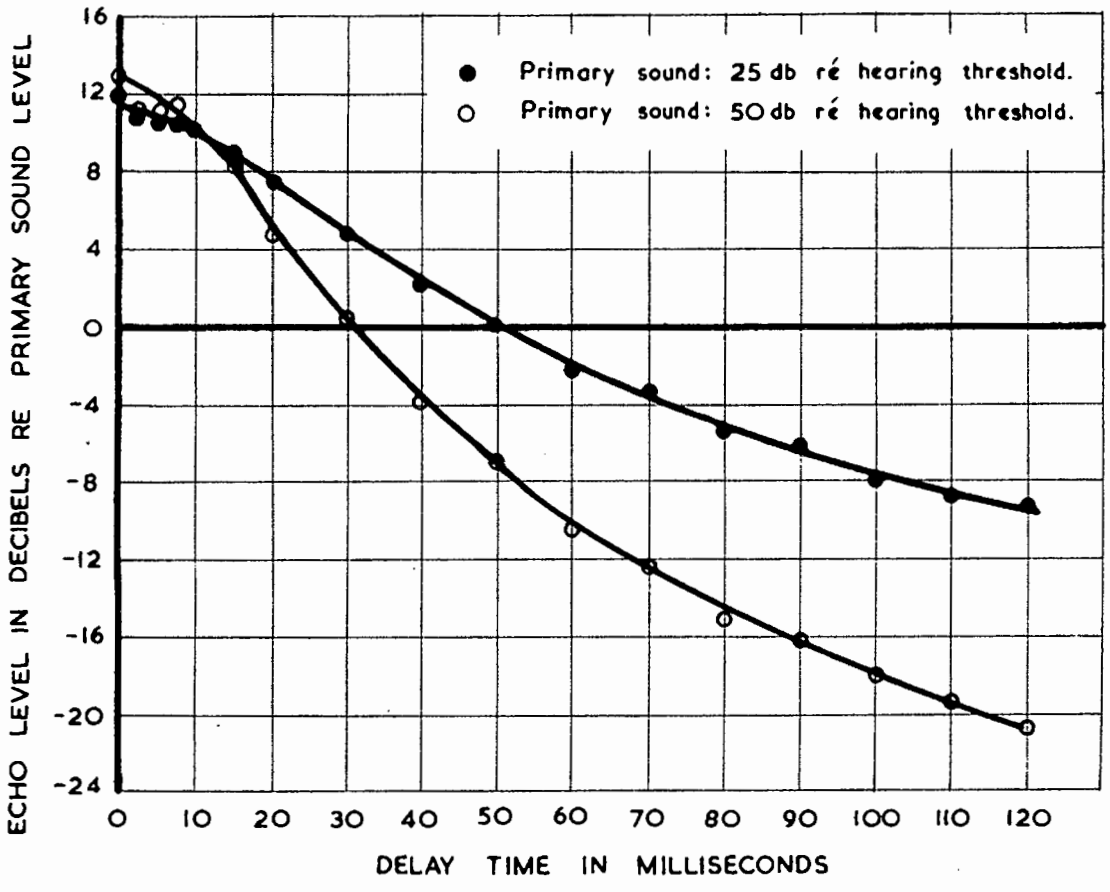
Points obtained with various observers showing the level in decibels re the primary sound required for an echo to become just audible.

FIG. 9



Points obtained with various observers showing the level in decibels re the primary sound required for an echo to become just audible.

FIG. 10



Curve showing the level in decibels ré the primary sound required by an echo to become just audible. The points shown are the mean of the values obtained by five observers.

FIG. 11

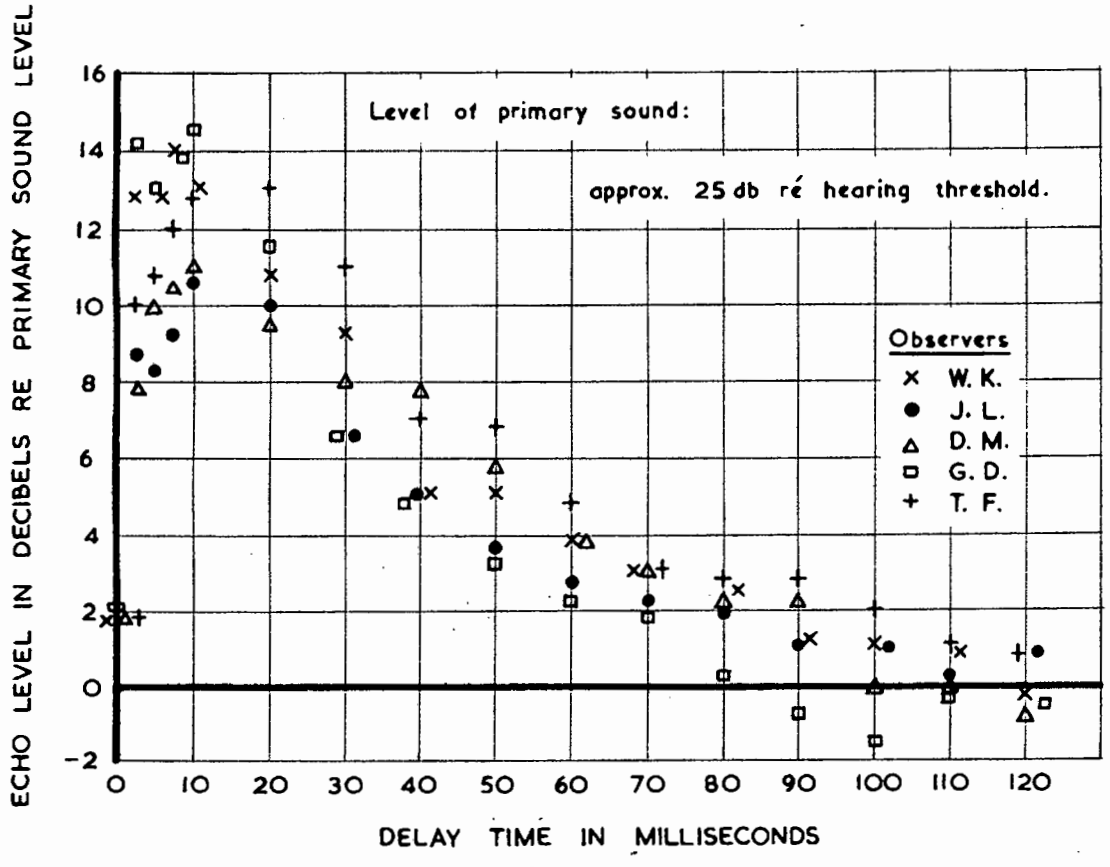


Figure showing the level in decibels ré that of the primary sound at which a single delayed echo sounds equal in loudness to primary sound.

FIG. 12

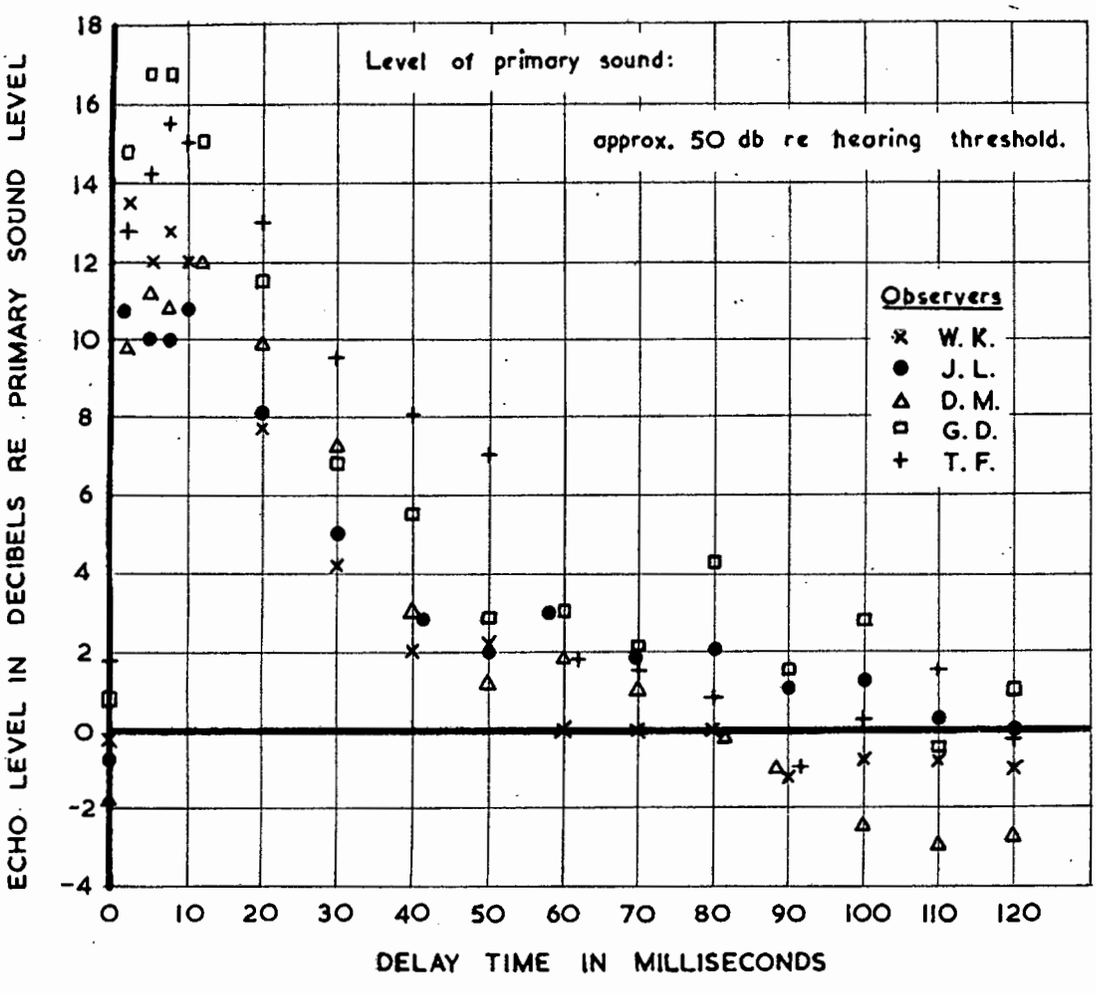
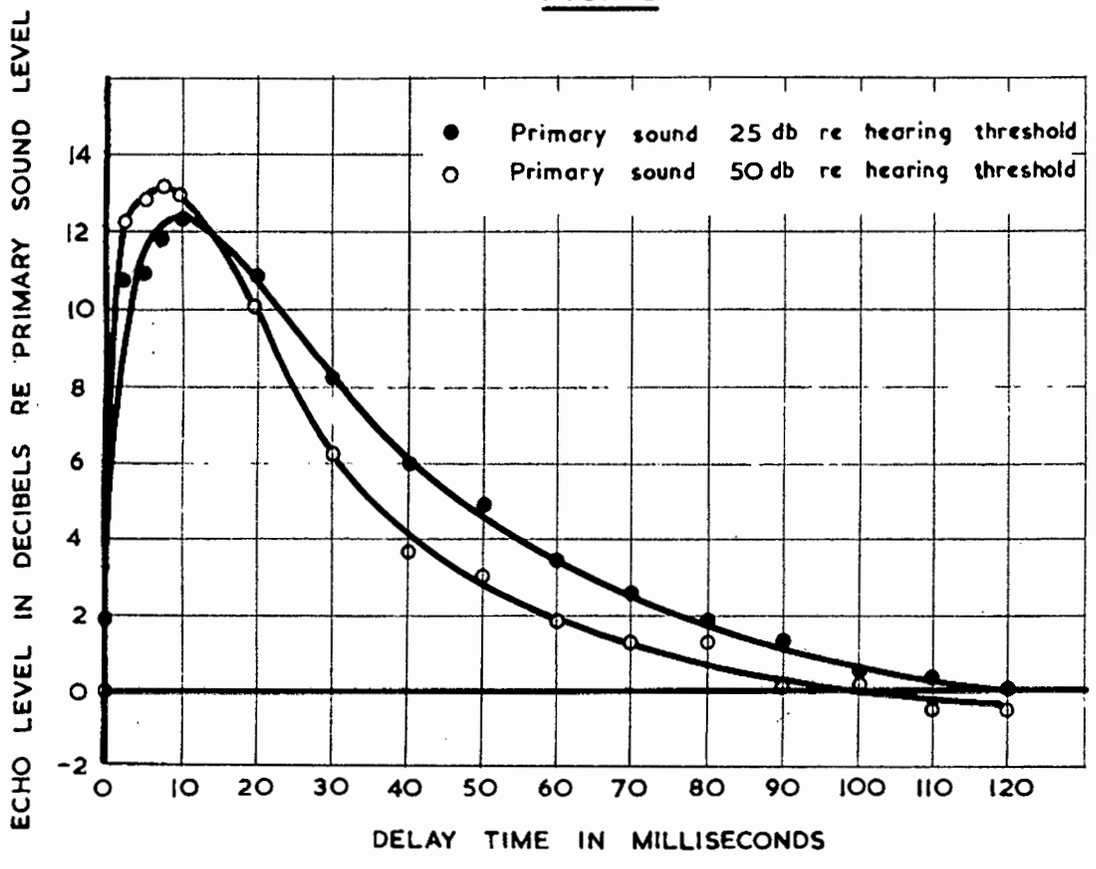


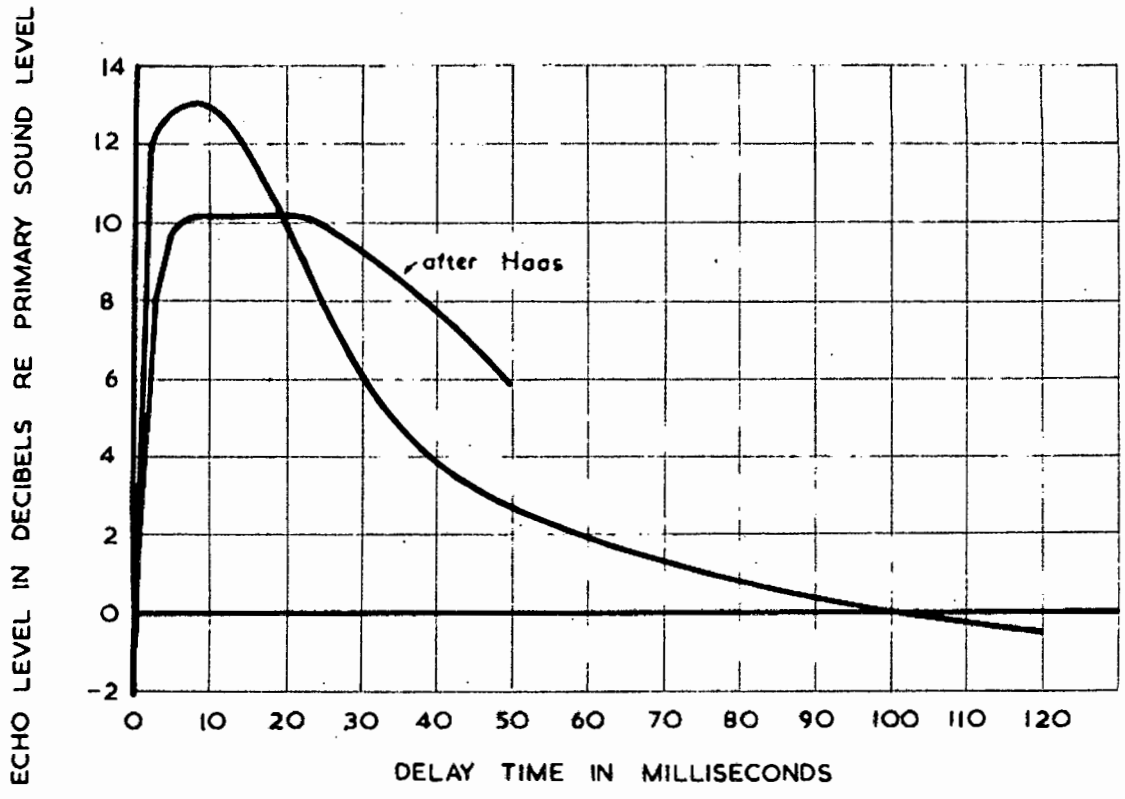
Figure showing the level in decibels re that of the primary sound, at which a single delayed echo sounds equal in loudness to the primary sound.

FIG. 13



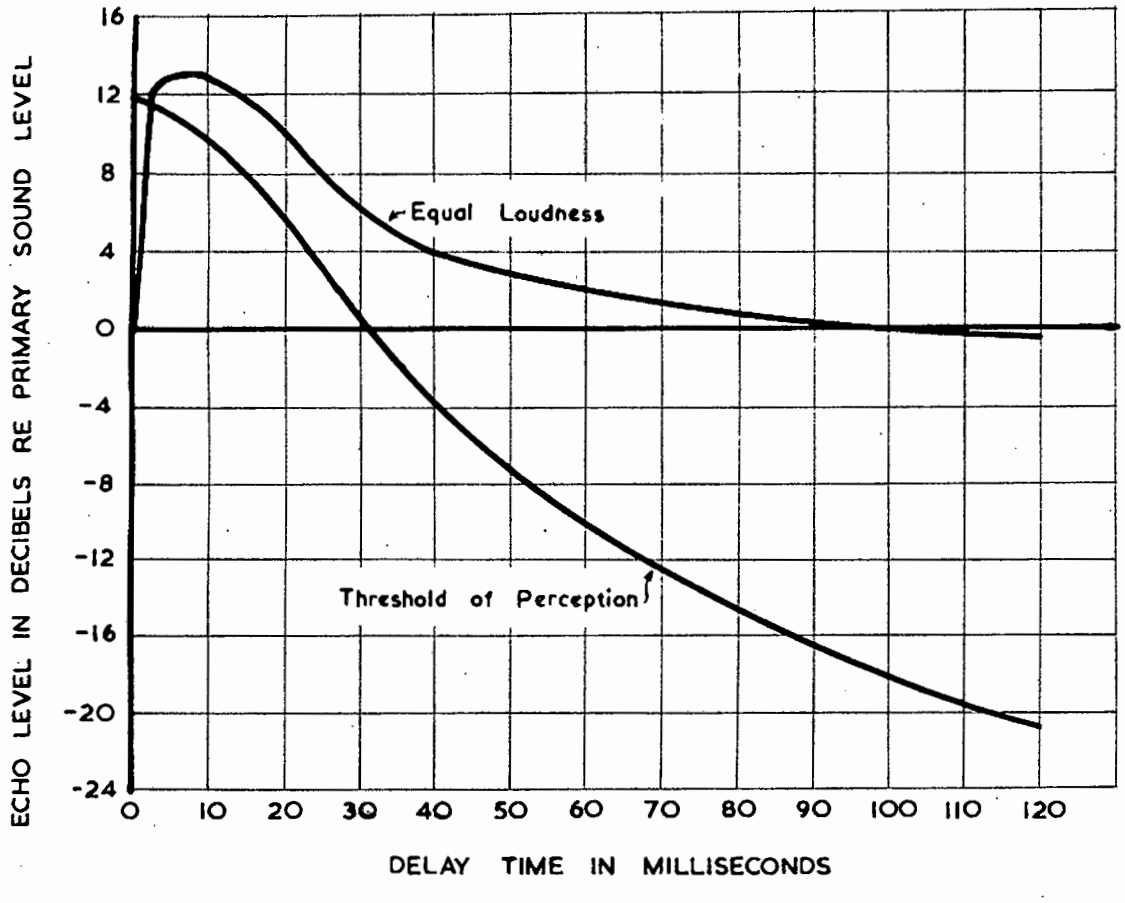
Curves showing the level required in decibels re the primary sound, for an echo to sound equal in loudness to the primary sound.

FIG. 14



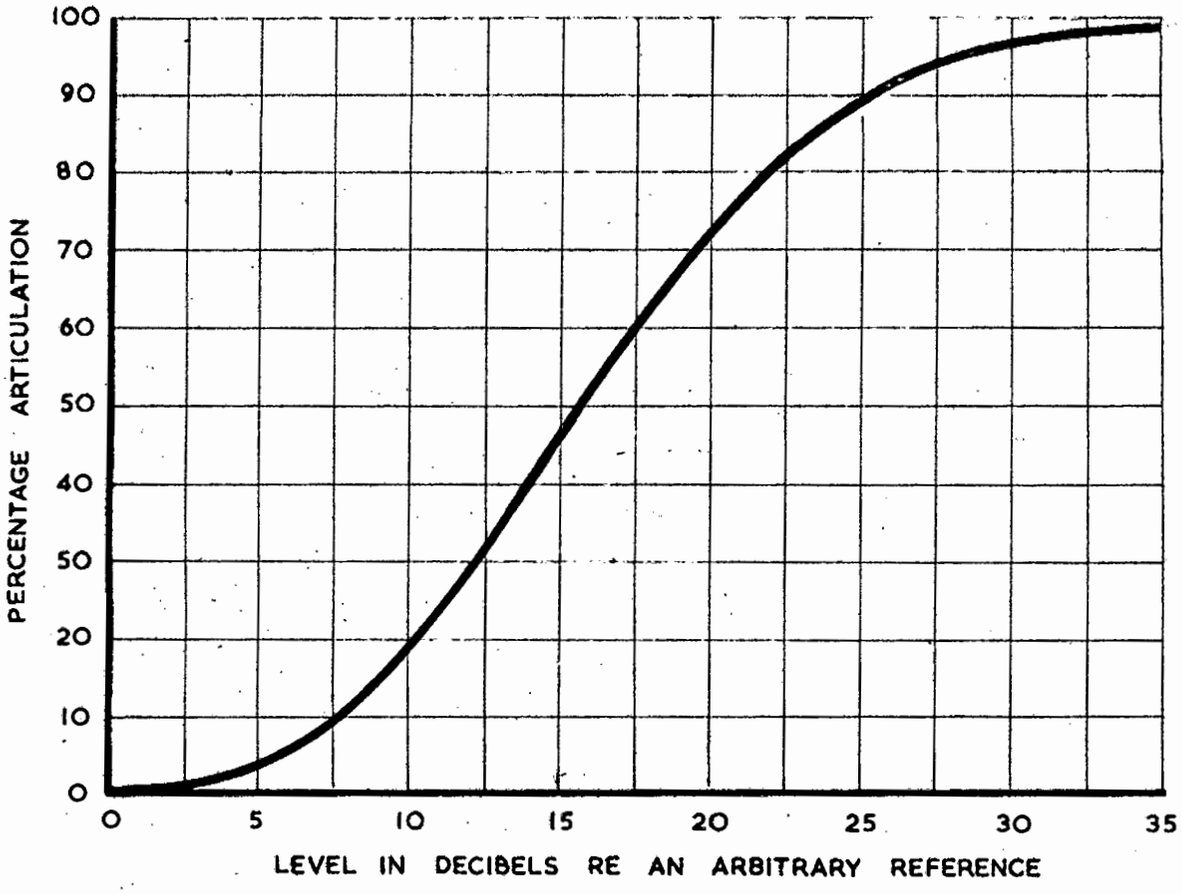
Comparison between equal loudness curve obtained in the anechoic chamber at a primary sound level of 50 db ré hearing threshold and that obtained by Haas.

FIG. 15



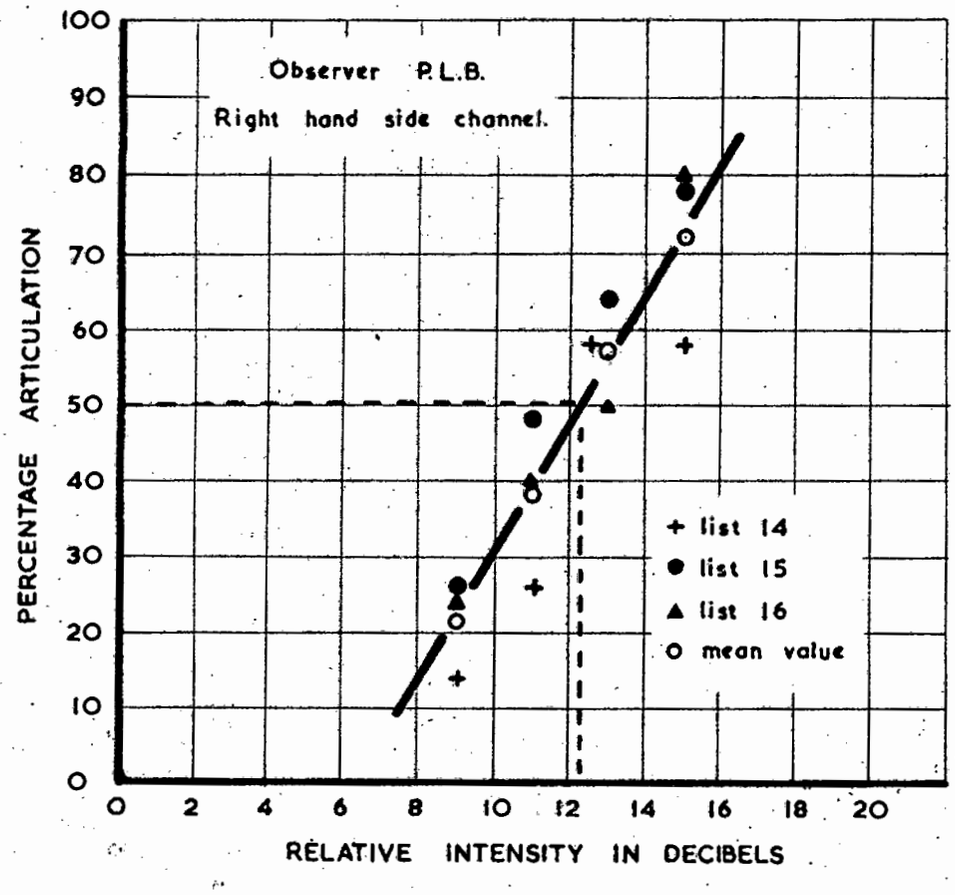
Curves showing equal loudness and threshold of perception contours obtained with primary sound levels of approximately 50 db ré hearing threshold.

FIG. 16



Curve showing relation between percentage articulation and level for a typical word list.

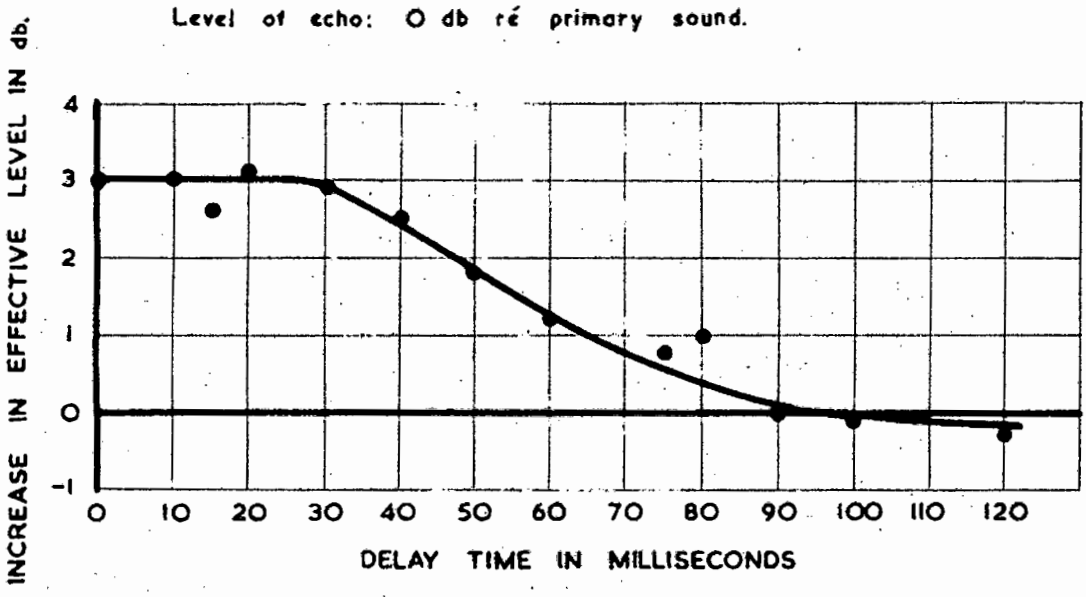
FIG. 17



Method of plotting articulation test results.

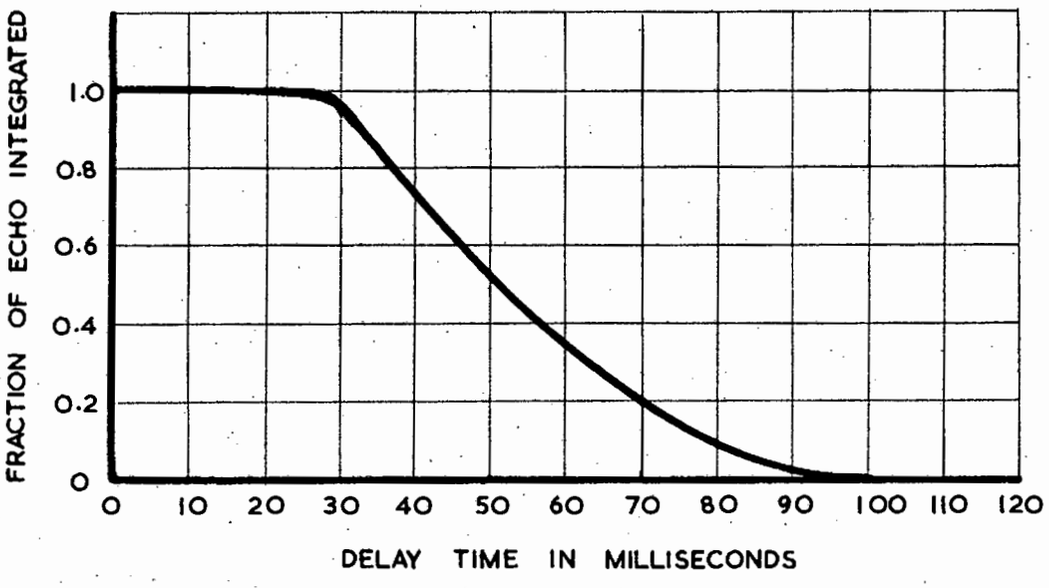
FIG. 18

Level of primary sound: approx. 25 db ré hearing threshold
Level of echo: 0 db ré primary sound.



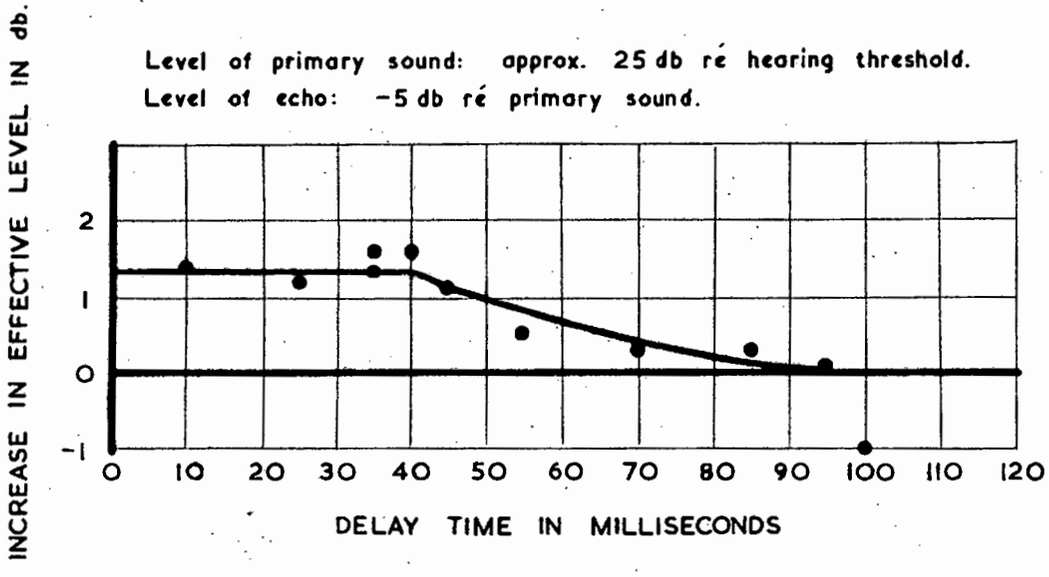
Curve showing the increase in effective level due to the presence of a single delayed echo.

FIG. 19



Curve showing the fraction integrated of an echo of level 0 db ré primary sound. Values were calculated from curve of fig. 19.

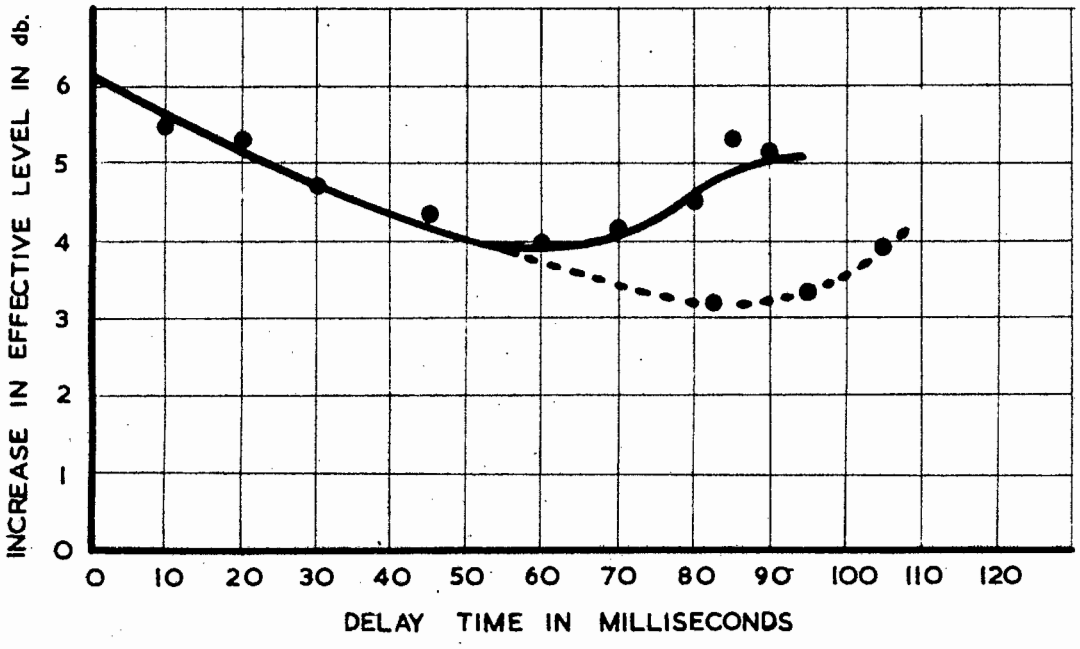
FIG. 20



Curve showing the increase in effective level due to the presence of a single delayed echo.

FIG. 21

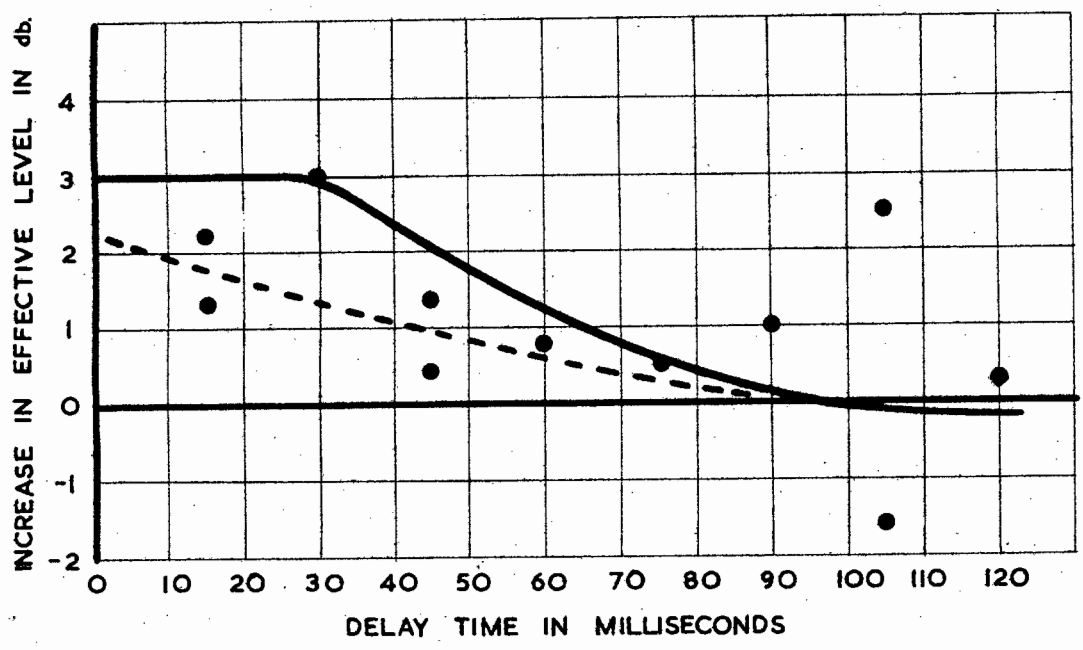
Level of primary sound: approx. 25 db ré hearing threshold.
Level of echo: +5 db ré primary sound.



Curve showing the increase in effective level due to the presence of a single delayed echo.

FIG. 22

Level of background random noise: approx. 50 db ré 0.0002 dynes/cm.^{2}
Level of primary sound: approx. 50 db ré hearing threshold.
Level of echo: 0 db ré primary sound.



Points showing the increase in effective level due to the presence of a single delayed echo in the presence of masking noise. Solid curve, shown for comparison, gives the results obtained in the absence of noise.

FIG. 23

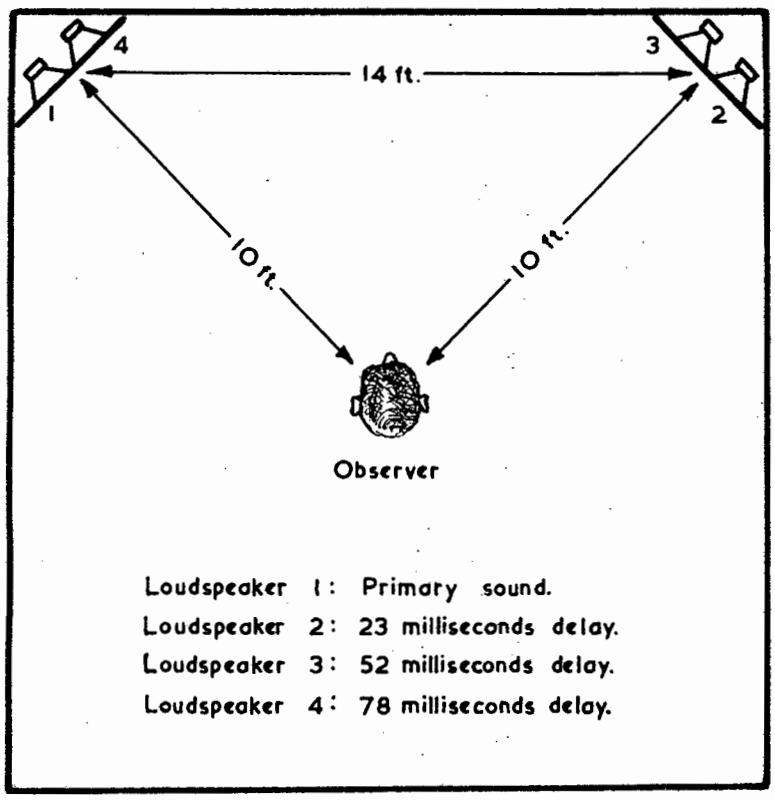
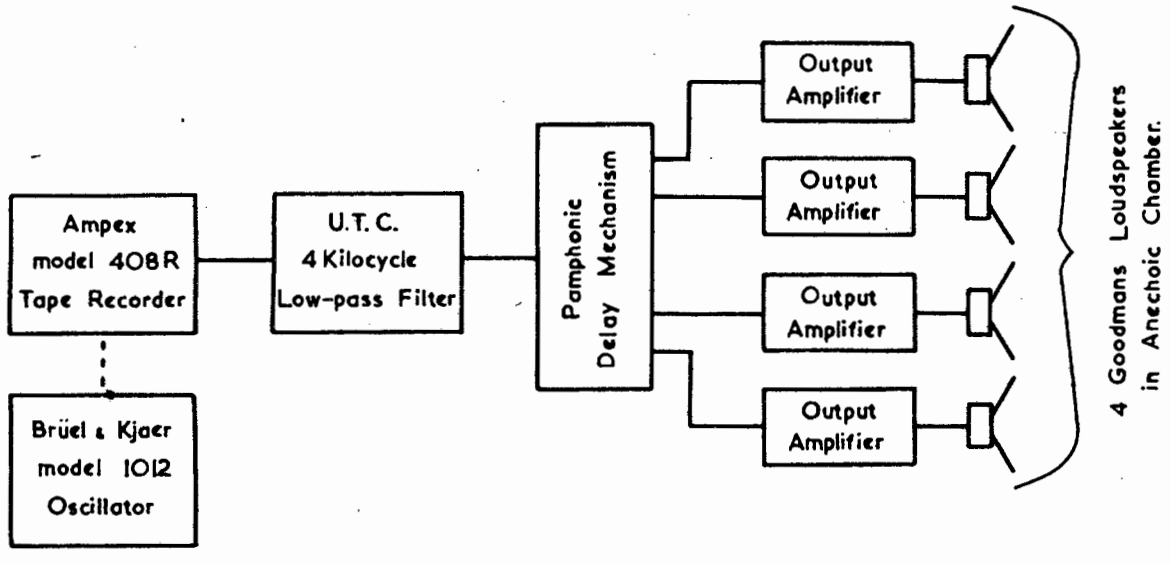


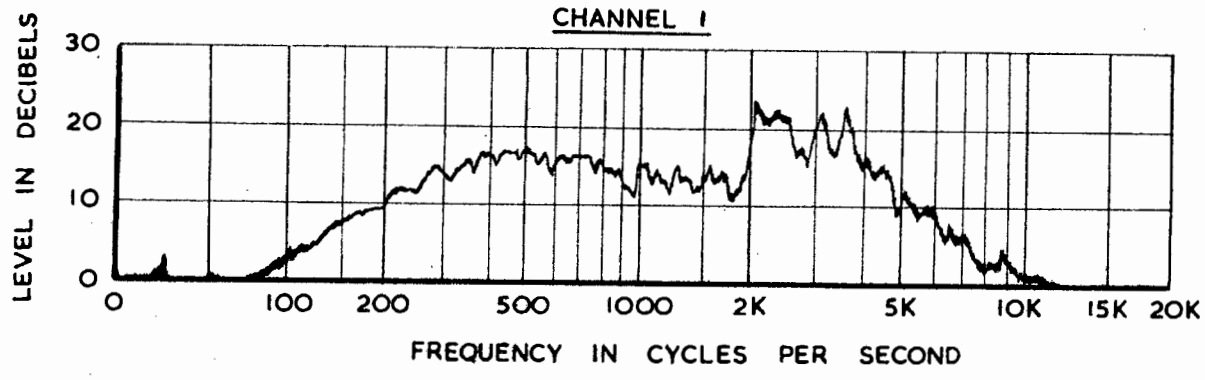
Diagram showing relative positions of observer and loudspeaker in anechoic chamber for tests with three echoes.

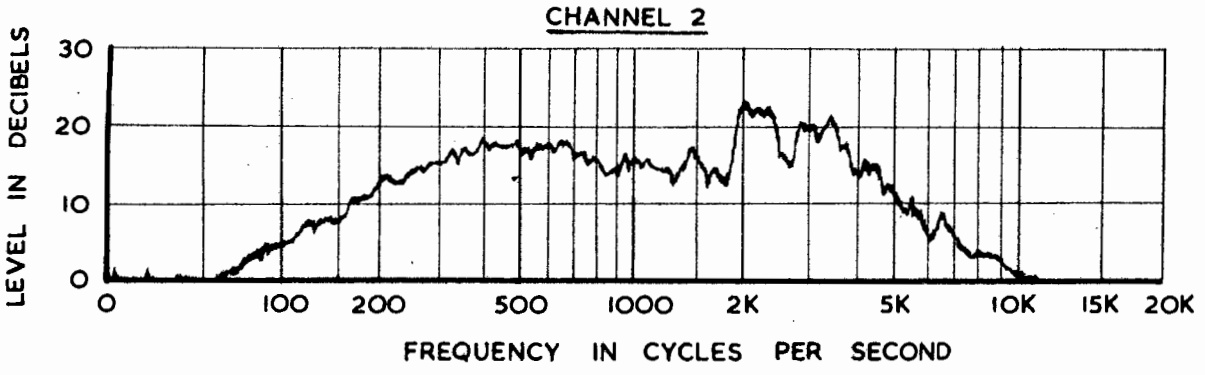
FIG. 24

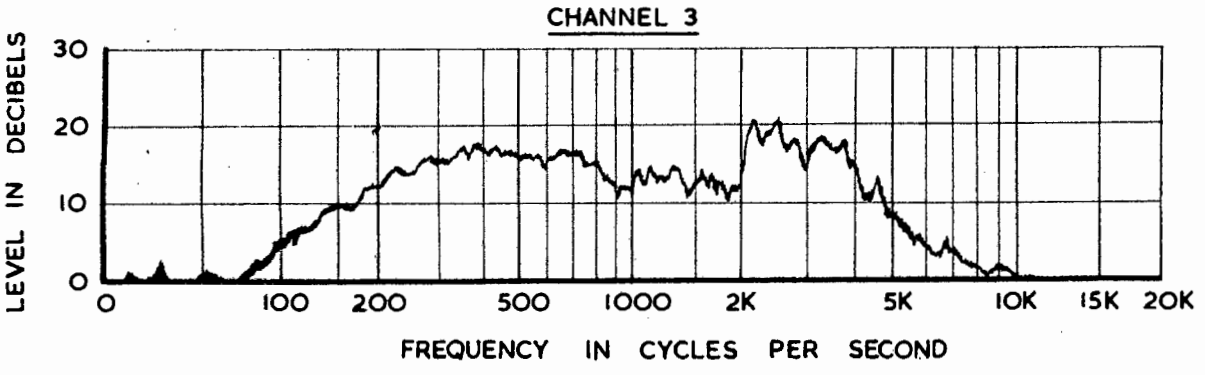


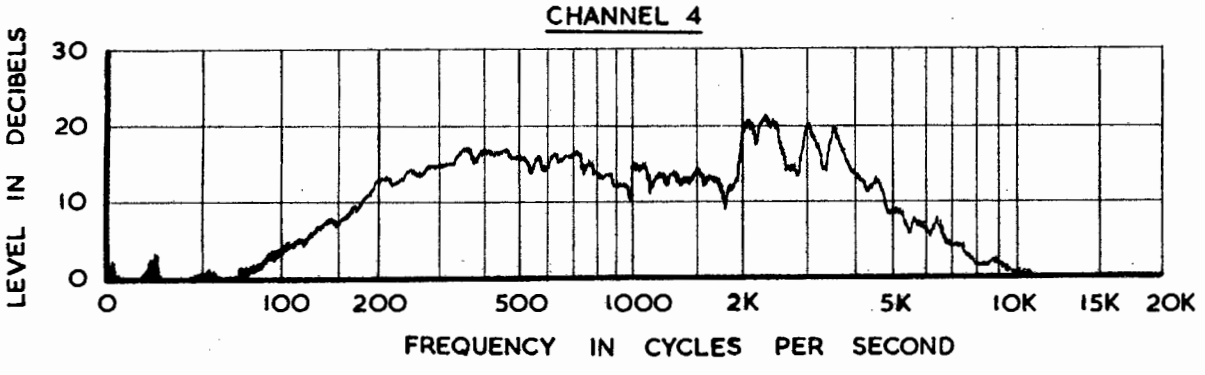
Block diagram of equipment used in tests with multiple echoes.

FIG. 25



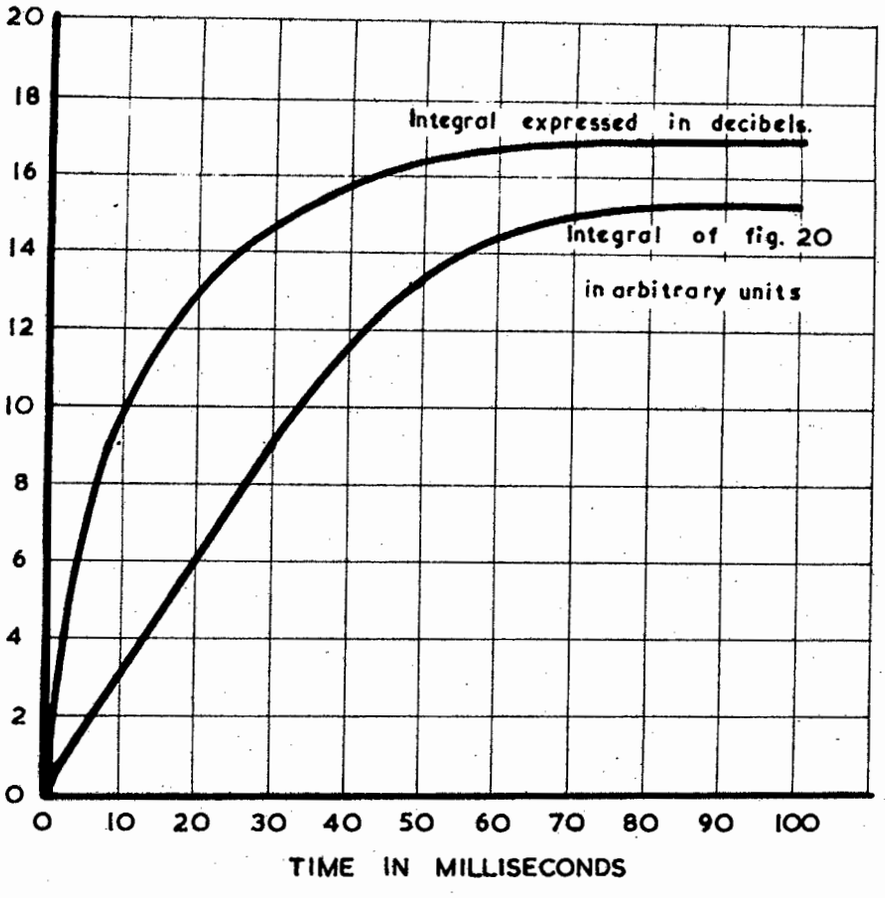






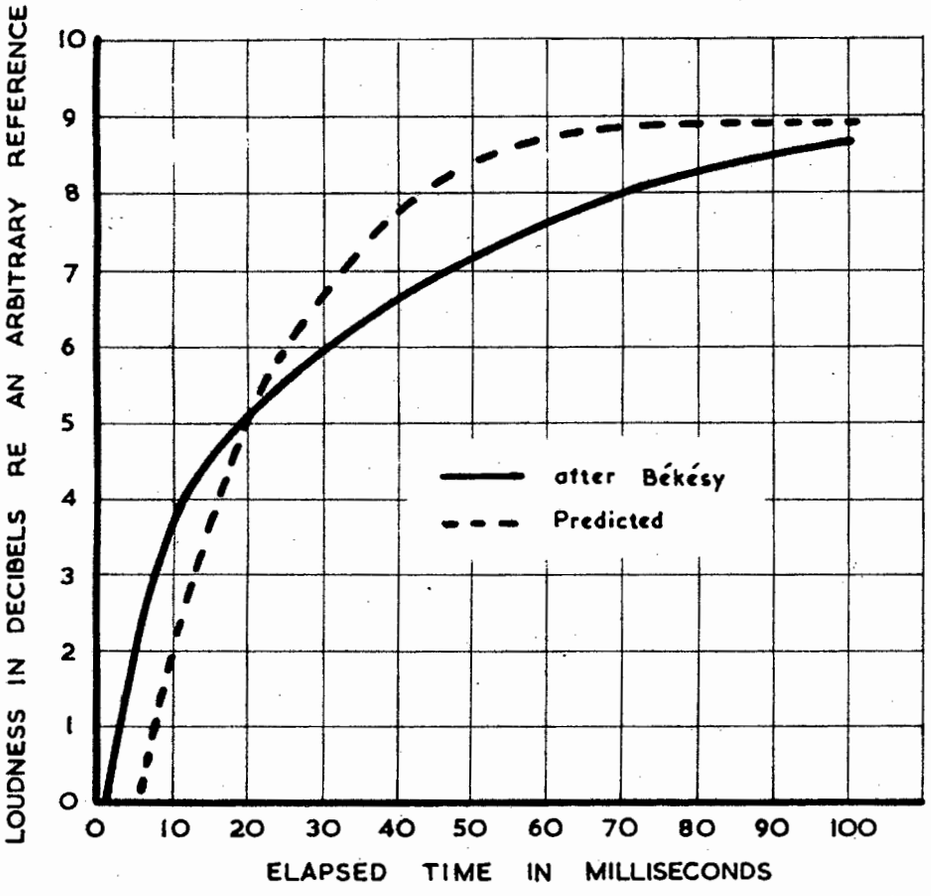
Overall frequency response of the four channels used in the tests with multiple echoes.

FIG. 26



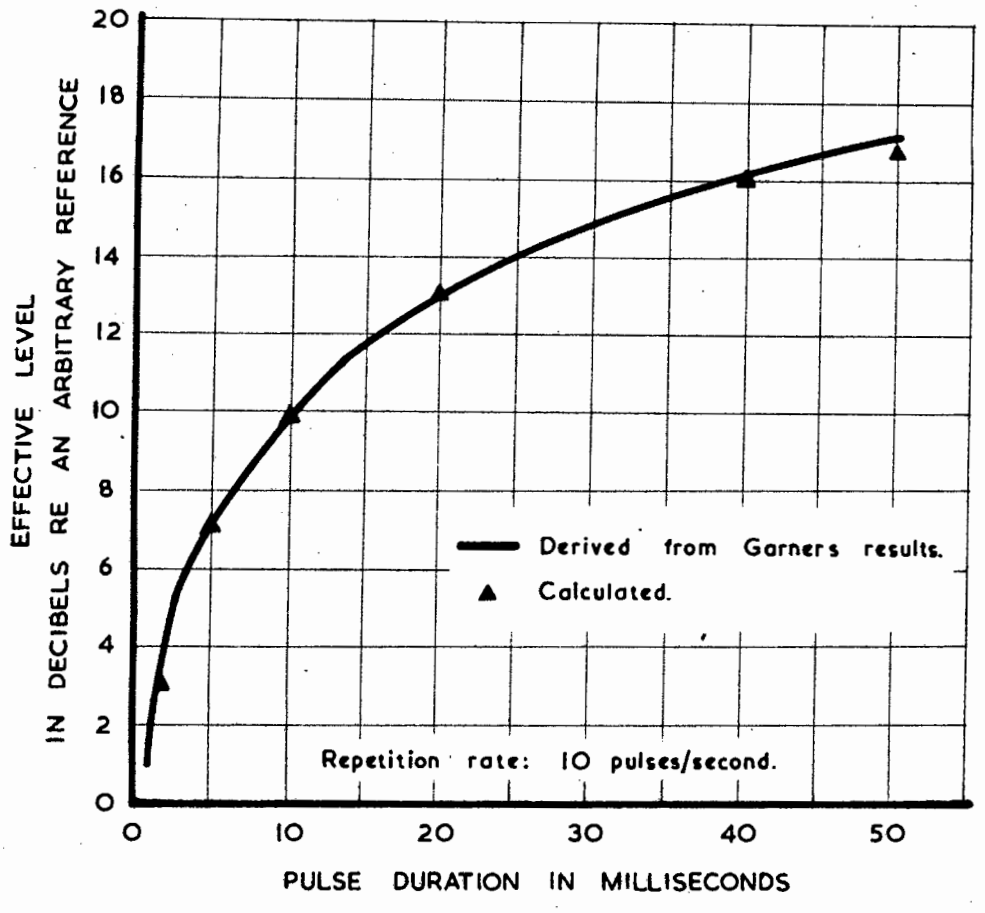
Curves showing the integral of fig. 20 expressed in arbitrary units, and in decibels ré an arbitrary reference.

FIG. 27



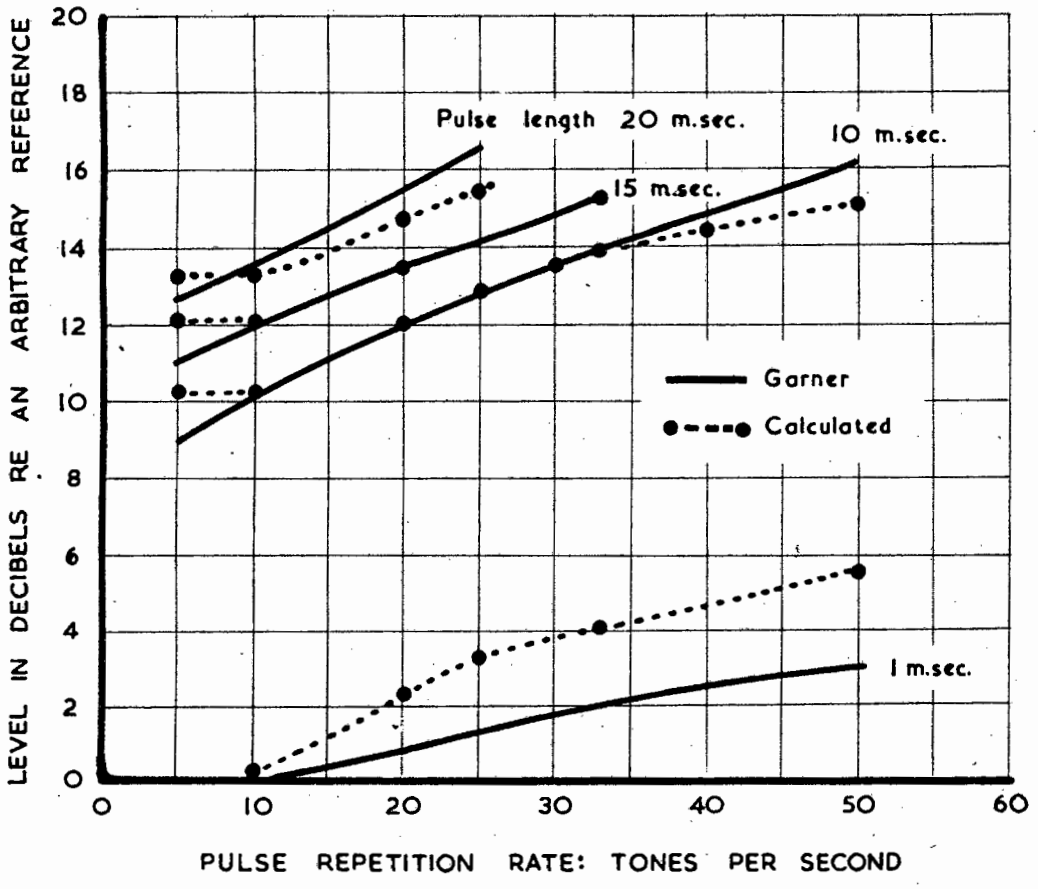
Curves showing the increase in loudness of a tone after switching on. The solid line represents measurements due to Békésy, the dotted line shows the predicted loudness increase based on the integration characteristics of the ear.

FIG. 28



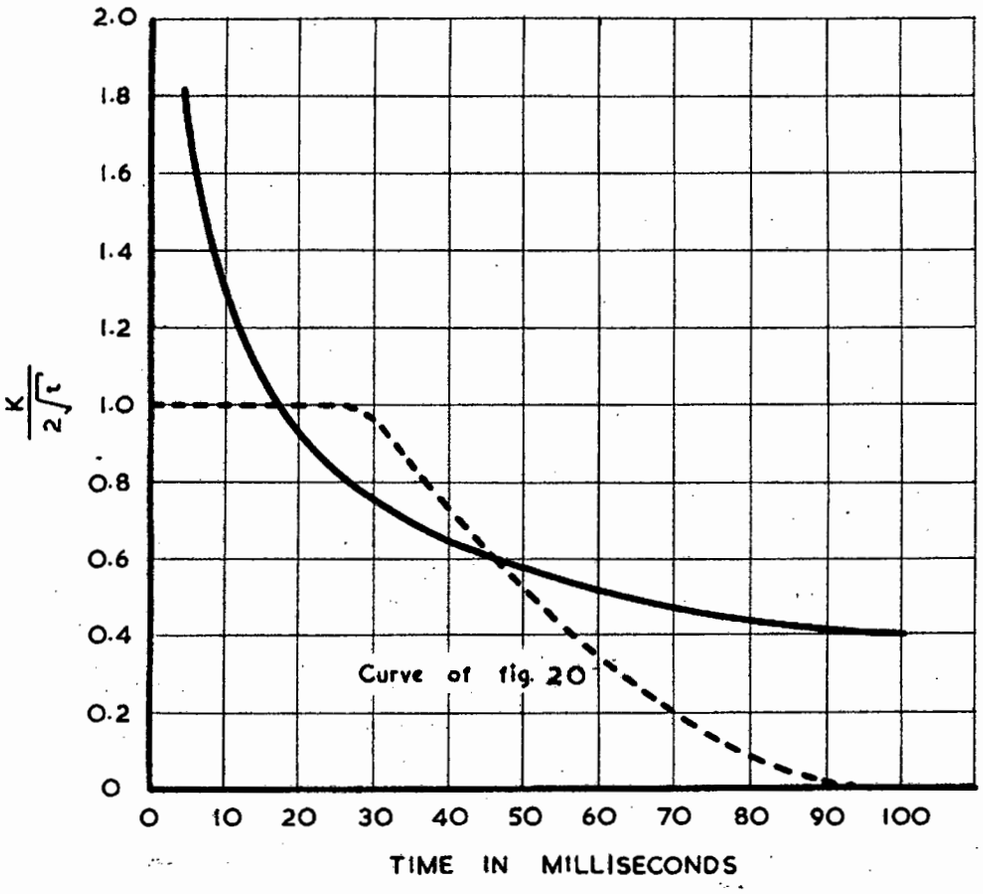
Curve showing the increase in effective level of keyed 1 kilocycle pulses as the pulse length is increased. (after Garner) The triangles show values calculated from the curve of fig. 27.

FIG. 29



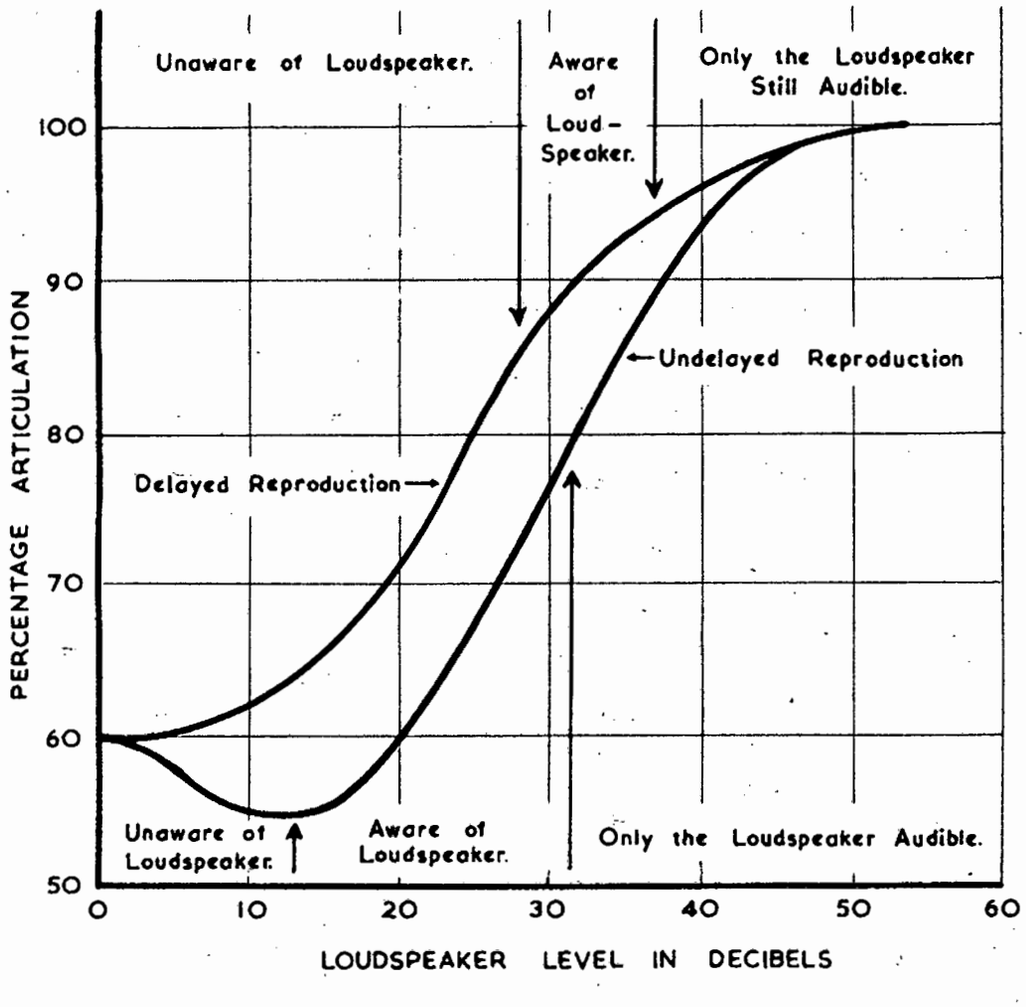
Curves showing the increase in effective level of pulses of various durations as the repetition rate is increased. (after Garner) The dotted lines show the values calculated from the integration characteristics of the ear.

FIG. 30



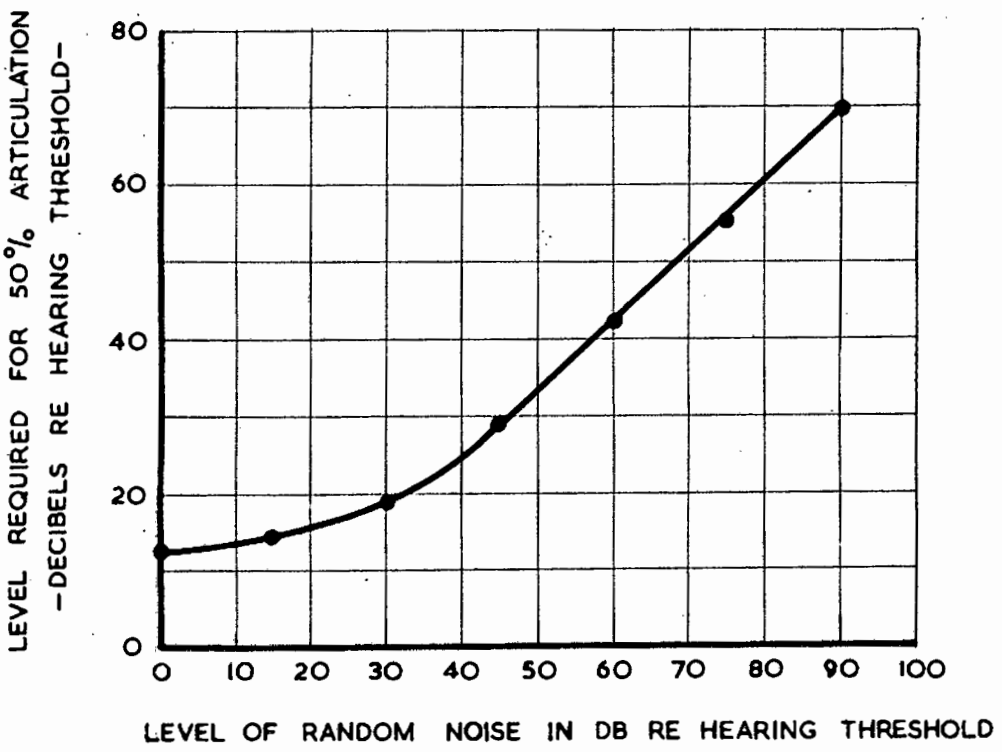
Curve showing the value of $\frac{K}{2\sqrt{t}}$ as a function of t.
K has been assigned an arbitrary value of 8.
Curve of fig. 20 is shown for comparison.

FIG. 33



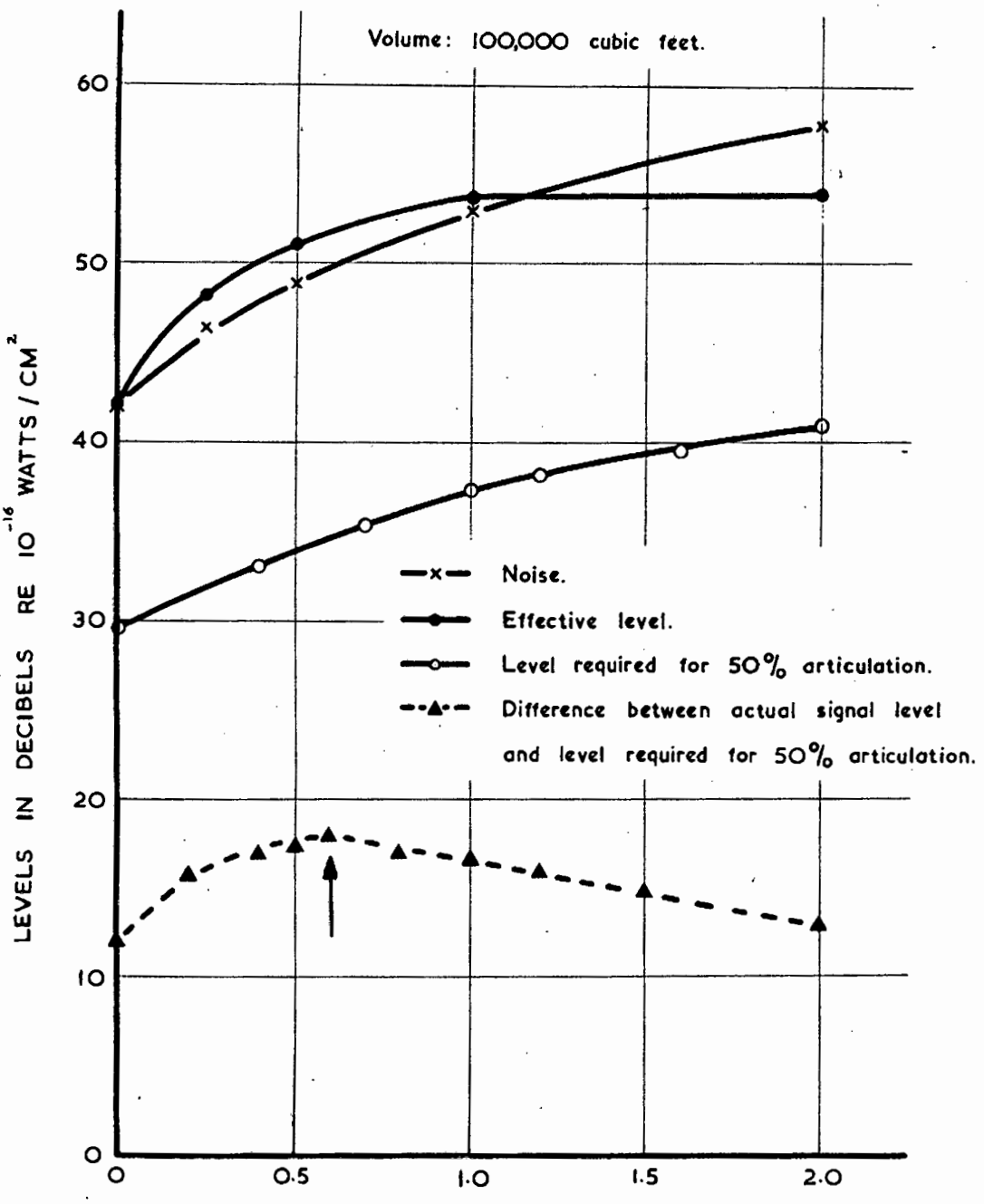
Articulation test results obtained with normal and delayed speech reinforcement in the House of Assembly. Cape Town.

FIG. 34



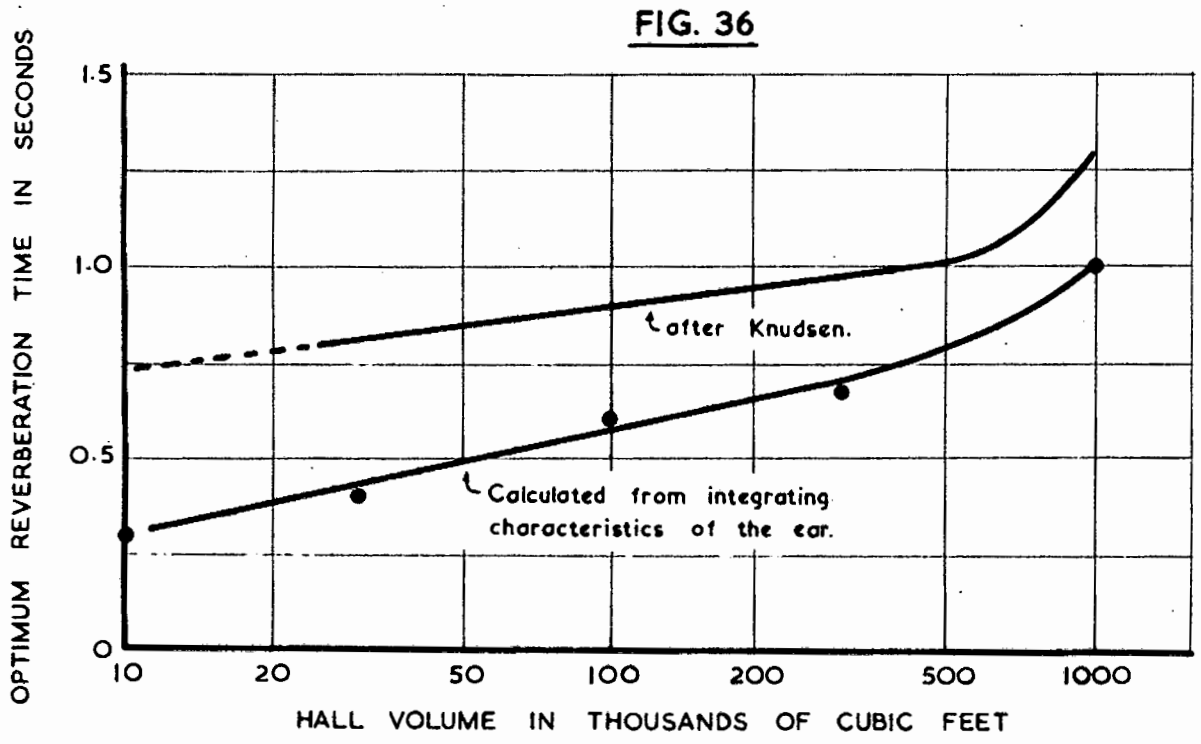
Curve showing signal level required in decibels ré hearing threshold to obtain a 50% articulation score in the presence of noise.

FIG. 35



Curves showing levels of effective signal and noise in a hall for different reverberation times.

FIG. 36



Curves showing optimum reverberation time as function of volume for halls intended for speach.

FIG. 37