An analysis of factors involved in musical ability, and the derivation of tests of musical aptitude

Davies, J. B.

AN ANALYSIS OF FACTORS INVOLVED IN MUSICAL ABILITY, AND THE DERIVATION OF TESTS OF MUSICAL APTITUDE.

J. B. Davies.

Thesis submitted for the degree of Ph.D.
October, 1969.

Department of Psychology, University of Durham.

ABSTRACT

Existing tests of musical ability or aptitude, with the exception of the Seashore battery, have largely been drawn up from the musician's standpoint. The use of musical material in the construction of test items has been almost universal; this approach being to a considerable extent the result of the view that only measurements taken 'in situ', that is in a musical context, have any value. This does not accord with the psychometric position which places emphasis on the predictive value of test items, but makes no statement about what the nature of items must be.

The present study is an attempt to construct a battery of tests of musical aptitude from the point of view of the psychologist. The approach is simply psychometric, and involves the use of elementary signal detection, and information, theory. The result is a battery of tests for use with schoolchildren between the ages of seven and eleven years. The data for reliability and validity, though by no means conclusive, suggests that such an approach is viable. In addition, there are reasons for supposing that the types of material selected have less experiential and cultural bias than other testing systems.

The work described falls into two main categories. Firstly there is an examination of existing test batteries, and of certain factorial studies of these batteries. In this section an attempt is made to define such terms as 'musical ability', 'aptitude', and 'ability'. Secondly, the construction of a test battery is described, together with certain

experiments performed. This leads on to the discussion of the results obtained as a result of administering the tests to some 2,000 schoolchildren. In terms of reliability the results are comparable with existing batteries. The indications from the validity data are also on the whole favourable, but further studies are desirable.

CONTENTS.

A tape of Test Form 'B' accompanies this thesis.

ACKNOWLEDGEMENTS.

The writer is indebted to the following people for help and assistance of all kinds in the preparation of this thesis:-

Professor F. V. Smith, for making available the facilities of this Department, and for general encouragement.
Mr. Douglas Graham, for supervising the work.
My wife, for general assistance at all stages of the work.
To the staff of all the schools who took part in the project.

Most important, to the children who worked the tests at all stages of development, for their unfailing good humour, energy, and tolerance.

Lastly, to the many other people, who by their help and encouragement have been instrumental in seeing this work to a conclusion.

Thanks are also due to the S.S.R.C. for providing the financial support for the project.

A list of all the schools and colleges who participated is given below:-

<u>Schools</u>: Gilesgate Junior Mixed.
Gilesgate Moor Junior Mixed.
St. Joseph's R.C., Durham.
Leamside Junior Mixed.
Sherburn Junior Mixed.
Durham Bluecoat Junior Mixed.
St. Hild's C. of E., Durham.
Belmont C. of E., Durham.
Durham Chorister School.
St. Margaret's C. of E., Durham.
Tow Law Junior Mixed.
Whinney Bank Junior Mixed, Middlesbrough.
Fairfield R. C., Stockton.

<u>Colleges</u>: The Northern School of Music, Manchester.
The Royal Manchester College of Music.
The Trinity College of Music.
The Guildhall School of Music and Drama.
Consett Technical College.
Durham University Education Department.

Sincere thanks are due to all the above.

INTRODUCTION.

The work described in the following pages falls broadly into two parts. The first part (involving sections 1 and 2) is mainly theoretical; the second (involving sections 3 to 5) is an account of certain experiments performed. Underlying this main division are the five sections themselves, the contents of which are as follows:-

Section 1 devotes considerable time to the consideration of factorial studies of tests of musical ability, and to problems associated with definition. Section 2 starts from a consideration of all the types of material available for use in a new test battery of musical ability, and considers and evaluates each type individually. Section 3 describes experiments leading to the production of a first, pilot, version of a new battery of tests of musical ability. In section 4, the development of the material contained in the pilot battery is described, and the changes necessary as a result of administering the tests to over 2,000 children are discussed.

Section 5 gives the conclusions from the study as a whole. These are fairly brief, and specific. Many accounts of the broader implications of musical research of this type have been given by other writers, and little purpose is served here by repetition.

Originally, the aim of the research was to produce a short battery of tests of musical aptitude for use in schools, which might have certain advantages in terms of freedom from certain cultural influences. As the work progressed the aims became wider, and the original intention was to some extent reformulated. A large number of test batteries for measuring various aspects of musical ability already exist, and the mere addition to the pile of one more battery seemed to some extent superfluous. Instead, it seemed more constructive to examine certain types of new test material, and to compare these with existing tests and measures. Final results would be regarded mainly as a guide for further research rather than as a fait accompli.

The general approach is psychological, and involves psychometrics, elementary signal detection, auditory measurement techniques, and other aspects of this discipline. This is in contrast to most other work in the field of measurement of musical aptitudes, where the approach has been basically musical. The end results of the two approaches are perhaps surprisingly similar, but there are different criteria underlying certain testing procedures. The principal point made by the work is that adequate measures of musical variables can be constructed without the adoption of the criteria used by musicians themselves. Whilst the approach adopted is in essence a simple 'job analysis', based on existing work and certain experiments performed, it is thought that the general line of investigation is not unsympathetic to the musicians viewpoint.

It is hoped that the research described, and the final version of the proposed test battery, might be of some assistance to teachers of music in junior schools; particularly when parents seek guidance on the purchase of musical instruments. Note, however, that nowhere is the intention to provide means by which children can be prevented from enjoying music as far as their abilities permit. The intention is solely to contribute, in some small way, to the advancement of music in education.

SECTION 1.

Musical Ability; the problem of definition.

Arnold Bentley suggests a possible definition of musical ability as "that characteristic, or those characteristics, which distinguish 'musical' persons from 'unmusical' persons", (Bentley, A., 1966). He acknowledges the difficulty of defining 'musical' as distinct from 'unmusical.' Seashore gives a more elaborate definition; "Musical talent is not one but a hierarchy of talents, branching out along certain trunk lines into the rich arborization, foliage, and fruitage of the tree, which we call the musical mind. The normal musical mind is first of all a normal mind. What makes it musical is the possession, in a serviceable degree, of those capacities which are essential for the hearing, the feeling, the understanding, and ordinarily, for some form of expression of music". Though a little more detailed, this is fundamentally the same as Bentley's definition, (Seashore, C. E., 1938).

On the other hand, Herbert Wing defined it operationally, in the following manner; "Many restrict the first term (musical ability) to the ability to play some musical instrument. But the teacher of music uses it in a wider sense that includes speed in learning to play, ability to perform the 'aural' tests discussed in the next chapter, and ability to carry out such musical activities as composing", (Wing, H., 1948).

The definitions of Bentley and Seashore are circular, since they define musical ability in terms of its own existence:- musical ability is an ability possessed by people who are said to possess it. We can make some kind of assessment of whether a person is musical or not by taking the consensus of opinion of a group of judges or experts, though such methods can be unreliable where personal bias or prejudice creeps in. If all that is required is the 'diagnosis' of musical ability, this type of definition will serve to some extent. However, if we wish to study individual aspects of this ability perhaps with a view to quantification or analysis, this kind of definition is no longer adequate. Viewed in this light the definition appears little more than a name or title for a group of undefined and as yet intangible phenomena.

On the other hand the type of definition offered by Wing attemps to define musical ability in terms of certain tasks which can be said to manifest the ability. The main difficulty here is that an almost endless list of tasks can be compiled, all of which can be said to 'have something to do with' musical ability, but which do not help in the search for a more basic, elemental, definition.

The problems of definition are not unlike those encountered in the literature on intelligence. Intelligence has been variously described as "the power of combining many separate impressions" (Spencer, 1895), or as the capacity for "the elaboration of a whole into its worth and meaning by means of combination, correction, and completion of numerous kindred relationships". (Ebbinghaus, 1897), or "the ability to select and maintain a definite psychic direction". (Binet, 1916) Definitions in terms of problem solving, "adaptation to new situations", ability to learn and many others have been put forward. To a great extent, the dilemmas apparent here are analogous to those encountered in attempting to define musical ability. Guilford, 1967, states "A definition that satisfies the needs of univocal communication must contain referents in the real world or must point unambiguously to something that points to referents in the real world". There are no such empirical referents for words such as "impressions", "whole", "worth", "psychic direction" and so on. The same applies, in the musical field, to words like "characteristic" and "musical mind".

The first truly operational definition of intelligence came from Boring, 1923, "...intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test". We can adopt a similar approach here and define musical ability as 'the capacity to do well in tests of musical ability'. We need to know, however, what it is that tests of musical ability or aptitude measure, if indeed they measure the same thing at all. Evidence on this point is by no means clear. One of the better investigations in this area was carried out by John McLeish,(1950). In this paper three different tests were examined; Standardised Tests of Musical Intelligence, by Herbert Wing; Measures of Musical Talent, by Carl Emil Seashore; and the Oregon Music Discrimination Test. Using a factor-analytic approach, McLeish claimed to have found the same common factor in all three tests. This common 'musical ability' factor he called Musical Cognition, and he defined it as "the ability to recognise and understand the nature of changes in musical or quasi-musical materials". This definition is important and raises several points which will be referred to later.

A definition of musical ability in terms of tests of musical ability (i.e. that musical ability is that which is measured by tests of musical ability) leads logically enough to the type of study carried out by McLeish. The following extract is from Guilford , and refers to intelligence, but the argument applies equally well in the present context of musical ability. "We are thus thrown completely on the tests for a definition of intelligence, and without proof that one intelligence test measures the same thing or things as another we have as many definitions of intelligence as there are different intelligence tests". Similarly, in the field of musical ability, we need proof that the existing tests are in fact measures of the same thing. The usual way of obtaining evidence on this point is to make factorial studies of tests which purport to or would appear to, measure certain common aspects of musical ability. Before proceeding further on this point, it is necessary to look at existing musical tests, in order to determine which appear to be tests of aptitude as distinct from tests of attainment.

Aptitude and Attainment.

The term 'musical ability', as used in much of the literature, does not distinguish between aptitude and attainment. Briefly, we can describe aptitude as a potential for certain types of development, regardless of experiential or environmental factors. Attainment on the other hand, is something that has been acquired as a result of the interaction between aptitude (or latent ability) and certain environmental conditions more or less conducive to the development of the potential behaviour.

Theoretically there is a clear-cut line separating tests of attainment from tests of natural aptitude, but in practice this is not so. Whilst it is fairly easy to devise tests of attainment, based on material that has been or should have been (according to some criterion) mastered, it is very difficult to devise aptitude tests that do not, to some extent, involve factors of experience or environment. Ideally, a test of aptitude should be such that anyone, regardless of past experience, approaches the test with precisely the same advantages and disadvantages as anyone else, apart from the constitutional factor to be measured. No such test has ever been devised; all we have are various approximations to this state of affairs in test batteries that involve greater or lesser degrees of experience.

At the outset, even though no formal definition has as yet been offered, it is necessary to redefine the term musical ability in such a way as to indicate its precise relationship with attainment and aptitude. Clearly, the definitions offered at the start of this section, in terms of 'something that musical people have', ability to play an instrument, ability to compose, or to perform certain musical tasks, and so on, have the two concepts inextricably bound up together. An easy way of clarifying the situation is to say that musical ability is something <u>that enables</u> a person to perform tasks of the type mentioned, with the implicit assumption that if a person cannot in fact perform any of these tasks, there may nevertheless be something in his make-up which would enable him to learn such tasks. It is in this sense that the majority of musicians use the term 'musical ability'. Teachers remark that a certain child "hasn't got an ounce of music in him", though they have consistently poured music into him for the preceding three years; the implication behind a remark of this kind is that the child in question is lacking something basic, and that this lack frustrates all attempts at musical attainment. On the other hand, Bell, F, (1928) reported that her son, Hugh Lowthian Bell, was 'full of music' at an age when he had no formal musical skills of any kind. It seems then that experience in the form of tuition, and even the acquisition of a certain level of skill, does not necessarily indicate the presence of musical ability; and also that its presence can be detected with confidence before any formalised musical training has taken place. Thus, although a psychological definition of 'ability' would involve both innate and environmental factors, it appears that the term 'musical ability' has been used to denote something much more like 'musical aptitude'. With this distinction in mind we can examine existing test material to try and determine what is orientated towards attainment, and what towards aptitude.

In some cases it is easy to see which tests are attainment tests and which are not. For instance, the Kwalwasser Test of Music Information and Appreciation can be seen to be, and was intended to be, an attainment test. In this battery, nine types of question material were used, involving: 1. Classification of artists. 2. Nationality of composers. 3. Composers of famous compositions. 4. Classification of composers. 5. General knowledge of composers and compositions. 6. Tone of orchestral instruments. 7. Classification of orchestral instruments. 8. General knowledge of instrumentation. 9. General knowledge of musical structure and form. On the other hand, the test battery of Carl Seashore was clearly intended to be a test of aptitude, involving tests of an almost psycho-physical nature concerning pitch, time, intensity, consonance, memory, rhythm. The position is less clear with some other batteries, however. The Wing Tests for example combine tests of appreciation with tests of aptitude. This in itself would present no difficulties, if the two types of test material were kept separate, but they are not. Scores for both types of item are lumped together and an overall total score is obtained, on the basis of which Wing suggests a formula for calculating 'musical age' (Approximate musical age = $\frac{\text{mark}-25}{3}$) and from this, a 'musical quotient', from the formula, based on I.Q.,

$$\frac{\text{musical age}}{\text{chronological age}} \times 100.$$

This lumping together of scores on all parts of the battery contrasts with Seashore's approach, in which the different sections of the test are represented separately, each on an individual histogram. Wing and Seashore represent opposing views in a dispute about the nature of musical ability. On the one hand Wing sees musical ability as being 'unitary' in nature; Seashore takes an 'atomistic' point of view. The 'unitary' theory states that musical ability is a single entity, and that the ability to perform musical tasks in general is dependent upon a common underlying ability. It is generally conceded by supporters of this theory, however, that this common underlying ability is in some way 'complex'. Wing, who worked with Cyril Burt, stresses the 'essential oneness' of musical ability, basing his evidence on inter-test correlations. A detailed account of a factorial study of inter-test correlations is given in the British Journal of Psychology,(1941, p, p. 341, Wing, H. D., A Factorial Study of Musical Tests.) Other workers who follow this viewpoint include Mursell, and more recently, McLeish, and Shuter. The opposite point of view is crystallised in Seashore's Measures of Musical Talent, which utilise a 'profile' system of scoring with each part of the test represented individually. The implication behind this approach is that high scores on one section of the test in no way imply high scores on other parts, and that there is no intrinsic reason why they should do so.

Evidence here is of an opposite nature to that adduced by the 'unitary' theorists, and studies have been performed showing that inter-test correlations are low. Supporters of this point of view include Spearman, Semeonoff, and more recently, Arnold Bentley. None of the evidence from either side is conclusive, however, and further evidence along the same lines is likely to be largely redundant, for the following reasons.

A study of the literature suggests that the kind of evidence produced is very closely tied to the type of tests used. Thus, evidence for the 'unitary' theory is drawn largely from test material based on the 'unitary' theory, and evidence for the 'specifics' or 'atomistic' argument tends to be drawn from tests based on the 'atomistic' approach. Whether the theory or the test came first is a chicken and egg argument; the point is that one only gets out of a test what is in it to start with.

The way in which different viewpoints lead directly to the production of test material of certain types can be seen clearly from the following extracts. One of the most formidable protagonists of the unitary approach was J. L. Mursell, who launched a scathing attack on the Seashore Tests, from the viewpoint of the 'unitary' school. On testing musical ability in general, he wrote: "Only the observations of the subject in various musical situations are a guide to the degree to which talent is present." (Mursell, J. L., 1937). On similar lines Lowery, H, (1932), writes, "If it is required to test for the presence of innate musical tendencies, the entire isolation of constituent factors in music is not likely to be of great service; rather ought a factor which is considered sufficiently worthy of special attention be brought into prominence with a musical background, the conditions of the testing being therefore analogous to those occurring in musical performance". It is interesting to note that Lowery devised a series of tests requiring subjects to make judgements about virtually intact musical structures, with no attempt to isolate separate component abilities. Subjects had to compare cadences and make judgements about 'completeness' or 'incompleteness', and there was also a series of tests of phrasing and of theme recognition. (Cadence, 1926. Also B. J. P., 1929, Musical Memory).

On the other hand, Seashore wrote of Mursell's approach, "It is my humble opinion that no creditable test of musical talent can be built on that theory". A defence of his point of view, as exemplified in his tests can be found in the Music Educators Journal, December, 1937. Spearman writes, on the subject of a general music factor, "Most of all perhaps, it might have been expected in the sphere of music, where not only innate instinct but also environmental encouragement are incomparably more favourable for some individuals than others. And yet just here the expected broad factor has been convincingly

disproved; the abilities to appreciate, for instance, the relation of pitch, loudness and rhythm have extremely low intercorrelations; no more, in fact, than must be attributed to G* alone". (Spearman, C, 1927). Semeonoff wrote, "The statistical treatment of the data suggests that it is impossible to postulate the existence of unitary musical ability".
Of the results found by Spearman, Herbert Wing wrote "Spearman's opinion was, however, in the main formed from results obtained with the Seashore tests, and his failure to find a group factor is, in my own opinion due to the doubtful nature of the tests used, for they did not involve appreciation in any marked degree". Semeonoff was also criticised because his tests were 'unmusical'.

The above arguments are in some cases confused and none is entirely convincing; though they illustrate the point that conclusions about the 'atomistic' or 'unitary' nature of musical ability are closely tied to evidence concerning the presence or absence of a common factor in the intercorrelations between certain types of testing material. It is also very clear that the results obtained are closely related to the type of test material used in the study. Thus, Spearman's results are derived from the psycho-physical tests of Seashore, which test more or

* For detailed explanation of G, see Spearman and Jones, W. Human Abilities, London, Macmillan, 1950.

less discrete physical capacities. Wing's results are based on his own test battery, in which variables are often confounded. It is clear, for instance, that he regarded as being of a 'doubtful nature' any test which did not involve 'appreciation in any marked degree'. It is difficult to reconcile this attitude with Wing's suggestion that tests 1 - 3 of his battery (which are not tests of appreciation) can be used in isolation from the rest of the test in certain circumstances. It is doubtful if further factorial studies along the same lines as those described, and using the same test material, will be of great help in suggesting any evidence which might clarify the situation, until certain fundamental issues are settled.

On the subject of intelligence tests, Guilford points out that some common referent or starting point is essential for all workers involved in test construction. "Such agreement could be achieved on a purely conventional basis as by popular vote or by the imposition of a constant test battery by some bureau of standards. A far better way of achieving unanimity of reference would be to find a foundation in psychological theory to which by experimental demonstrations those who construct tests would feel persuaded to assent". In the field of testing musical ability, no such foundation exists. In fact the greatest emphasis has been, perhaps, on test production with 'experimental demonstration' taking a second place.

In this sense, much of the work by such writers as Stumpf, Kohler, and Guttman seems to have been lost sight of.

Evidence from Factor Analysis.

One of the major problems arises from the tendency of workers in this field to take data obtained from a particular approach to testing, and then to imply that results can be generalised to include other approaches. For example, a comprehensive factorial study was performed by Herbert Wing, on his own tests. This consisted of three factor analyses, one in 1936, and a main and a subsidiary study in 1941. Common factors emerged from all the studies, though whether the different studies produced the same factors is by no means clear. Wing's conclusion that there is a "general ability to perceive and appreciate music" must be interpreted with caution; his results certainly tell us something about the structure of his own tests, but do not necessarily tell us anything about the construct of musical ability, nor about other test batteries. Similarly, the finding that another test battery also yields a common factor suggests nothing, unless we can show that the factor in question is identifiable as the same one in both batteries. Finally, if a factor emerges, which is identifiable in several batteries, it must be shown that the factor in question is pertinent to the construct in question, and not merely an irrelevant by-product, or an artefact of measurement techniques. It has already been stated that several variables appear to be present in the Wing battery; for instance, tests of aptitude and tests of appreciation* appear side by side; in the memory test, as will be shown later, melodic memory and pitch discrimination are intermingled; and there are several others. In a re-analysis by Holmstrom, 1963, seven significant factors were found, but at the end of this no meaningful interpretation was possible as residuals were still too high. However, another study by Faulds (1959) found factors, using the Wing tests, that were comparable with those found earlier by Wing. On the other hand, Spearman's study of the Seashore Test battery produced a series of low inter-test correlations. Though the results were specific to the one test, the conclusion implied things about musical ability in general, namely that the existence of a general factor for musical ability had been "convincingly disproved".

* The fact that tests of appreciation and of aptitude occur side by side would not in itself matter if both were measuring the same thing. There are reasons to suppose that they are not, however. Scores on tests of appreciation are far more likely to be a function of environment or experience than scores on certain other types of test material, and it has already been suggested that musical ability, defined in terms of '<u>aptitude</u>', is not a function of environment or experience. (see p. 5).

A comprehensive summary of the main factorial studies of musical ability is included in Shuter's book, appendix II. Unfortunately, very few of the studies described here compare complete batteries with other complete batteries. Instead, the tendency is to compare selected items from various batteries. In this respect McLeish's study is the only one comparing complete batteries; his finding of a common factor, underlying batteries as different in conception as the Wing and the Seashore, is thus doubly impressive. The identification of common factors underlying complete and undoctored test batteries is likely to be of more fundamental importance than similar results obtained from selected sub-tests, where initial choice might select for a specific factor; only the former can provide convincing evidence for a group factor theory of musical ability. In certain studies it would also be most useful if the actual intercorrelations between the sub-tests of the test examined were included, since it is possible to obtain high factor loadings even if the general degree of relationship between the parts of the test is fairly low. Finally, on the topic of factorial studies in general, the interpretation of results is often very difficult in this field. On interpretation, Anastasi writes, "Once the ..... factor matrix has been computed, we can proceed with the interpretation and naming of factors.

This step calls for psychological insight rather than statistical training. To learn the nature of a particular factor, we simply examine the tests having high loadings on that factor and we try to discover what psychological process they have in common. The more tests there are with high loadings on a given factor, the more clearly we can define the nature of the factor". In the field of musical ability, however, the data from factor analyses is by no means clearly defined, with the result that interpretation is rather more difficult. Below is given a brief list of some selected analyses, in which it is difficult to see a common 'psychological process':-

Karlin, 1941. 1st factor:- Drake Memory. 732, Retentivity. 625, Intervals. 564, Pitch. 448, Rhythm.443.

Holmstrom, 1963. 1st factor:- Pitch. 59, Memory. 61, Rhythm. 54 and .53, Music marks. 46, Chords. 32 or .28, Intelligence. 27.

Shuter, 1964. 1st factor:- Chords. 753, Memory. 569, Pitch. 470, Intensity. 466, Phrasing. 388, Harmony. 296.

It is important to bear in mind, when examining data of this kind, that one person's concept of a pitch test, or a harmony test, is not necessarily another's. For example, two sub-tests from different test batteries might both bear the same name, such as 'Pitch' or 'Memory'. However, the sub-tests might be quite different in nature, and may in fact measure different things. Similarly, high factor loadings on tests from different batteries are difficult to interpret where the comparable loadings have been obtained in different factorial studies. Though the loadings are comparable, and the sub-tests have similar titles, the factor loadings might have entirely different meanings.

One point does emerge from the factorial studies, however. There is a tendency for tests of 'memory', even though these are by no means of uniform type, to have high factor loadings on the largest factor. This is important for the present study, as memory is seen as being an important factor in all perceptual tests of musical ability.

The Study by John McLeish.

The difficulty of infering psychological processes from the factor analyses performed has led to the formulation of several more or less unsatisfactory definitions as to what the common factor might be. It has been variously described as 'musicality' (Holmstrom), 'music factor' (Faulds), 'musical cognition' (McLeish), and simply 'music' (Holmstrom). Several writers refer to it simply as a 'general' or 'broad' factor without further definition. (Wing, Vernon) In the study by John McLeish, 1968, an attempt is made to suggest what in fact the common factor might be. This definition is perhaps the only one which suggests in an explicit manner what the underlying psychological processes are. Musical cognition is defined as "the ability to recognise and understand the nature of changes in musical or quasi-musical materials", followed by the comment, "The size and nature of the intercorrelations are such that we may be certain that if Wing measures musical ability so does Seashore". (The last sentence seems to be a cautionary one, implying that even if we can identify the same factor in the two batteries, it is possible that the factor is not a 'music factor' but something arising from similarities in the nature of the test material itself, and specific to the testing situation.) The two crucial words in McLeish's definition are 'recognise' and 'understand'. It is by no means clear what is meant by 'understanding' in the field of music, and no real attempt will be made here to clarify this problem. The word 'recognise' however has fairly specific connotations in the psychological literature. Specifically, recognition is used as a yardstick for retention; and it particularly implies the ability to pick out certain stimuli in the present situation on the basis of past stimulation or experience. Recognition is in fact a type of memory.

Below is given a table of the results found by McLeish in the study quoted:-

	Tests analysed separately.			Tests together.
	First Factor.	Intelligence.	First Factor. minus Intelligence.	First Factor.
I Seashore Tests.				
Pitch	.62	.22	.59	.66
Intensity	.48	.13	.46	.48
Consonance	.39	.03	.43	.47
Memory	.87	.32	.87	.82
Time	.42	.13	.39	.30
Rhythm	.28	.17	.27	.37
% of total variance	29.0%	3.6%	25.8%	
II. Wing Tests.				
Chord Analysis	.70	.13	.71	.65
Pitch Change	.76	.23	.76	.78
Memory	.76	.25	.76	.78
Rhythmic Accent	.54	.30	.50	.45
Harmony	.66	.39	.60	.66
Intensity	.52	.34	.34	.52
Phrasing	.46	.23	.39	.41
% of total variance	45.0%	7.75%	37.3%	
III Oregon Test.				
Appreciation				.77
Nature of Change				.66
Percentage of total variance				38.8%

The frequency with which tests of memory consistently achieve high factor loadings has already been mentioned. A few workers have actually defined their main factors in terms of some form of memory. For instance Drake (1939) chose to call his main factor 'Memory'; and Karlin (1941) used the term 'Memory for Form'. In the above table the high loadings for memory are again observable. On the Seashore Tests, the memory sub-test has the highest loading of all, when analysed both separately and in combination. The second highest loading is for pitch. On the Wing Tests pitch and memory are joint highest in factor loading. Here again, it is very hard to see what psychological process underlies tests of memory and of pitch, to account for these loadings.

The conclusions that McLeish drew from his study were briefly as follows:-

1. "That Seashore was in error in that musical talents are NOT specific in their nature, in other words there is a group factor of musical ability.

2. That Burt was in error inasmuch as Wing's piano tests, using musical stimuli, measure precisely the same factors as do Seashore's laboratory-contrived psycho-physical 'measures'.

3. That in the structure of musical cognition, musical memory, pitch discrimination and ability to analyse chords have greater weight than the ability to discriminate differences in rhythm, time and intensity. "

In general one would agree with these conclusions, with the following reservations:-

Firstly, though Seashore's theory of complete specificity has been disproved by the finding of a group factor of musical ability, there is still considerable specificity in the tests. Secondly, though pitch discrimination and musical memory have considerable weight in both test batteries, the conclusion that chord analysis is also a major component cannot be substantiated since neither version of the Seashore Tests contained any chord analysis material.

The Analogy with Signal Detection.

The prominent position occupied by memory and pitch in this study is duplicated in many other studies. A brief description of the nature of the test material used in certain studies may help to throw some light on why this should be. Bearing in mind the high loadings often found on memory tests, and specifically the high loadings found in the McLeish study, we can examine the material used in the tests in the light of the literature on memory, and especially short term memory. Basically, there is no difference between the 'memory' tests of Seashore and Wing. Both involve the subject in listening to a short tune; the tune is then repeated, but with one note altered, and the subject is asked to indicate which note in the sequence was different in the second playing. There are two minor differences in procedure, namely, Wing used a number of items in which the two renditions were the same; also, whilst Wing used overtly 'melodic' material, Seashore's material was "purposely selected to form no melodic line."* Viewed as tasks in short term memory, the two memory tests are virtually identical, except that they are not matched for codability or length. The processes leading up to a correct response are as follows:-

> The subject hears the first rendition, and holds it in short term memory. He hears the second rendition and compares each constituent tone, as it arrives, with his 'copy'. When we hears a tone which is discrepant with the copy tone he indicates that that particular tone has been changed, by marking on his paper. In fact, in making his answer the subject has compared two tones, one which he hears, and one which is a memory 'copy' and made a decision whether they are 'same' or 'different'.

* There are reasons for doubting the implied clear-cut distinction between melodic and non-melodic. It seems more probable that there is simply a continuum of codability. Gardner, S, writes, "There are no unrelated tones in music." (School for Violin Study Based on Harmonic Thinking, Carl Fischer, Inc. N.Y. 1939).

We can represent this situation diagramatically, as follows:-

s 1	s 3	s 6	s 7	s 9	D 3	s 6	s 4	s 8	First Tune.
s 1	s 3	s 6	s 7	s 9	D 4	s 6	s 4	s 8	Second Tune.

In the diagram, numbers indicate tonal elements of the tune in two successive presentations. The letters 's' and 'D' indicate whether elements are 'same' or 'different'. If a subject correctly selects the two shaded elements as being different, he has made a successful pitch discrimination, and the remaining material serves simply as interference. An analogy with signal detection theory presents itself, since we can describe the distribution of 'same' and 'different' elements in the form of a signal to noise ratio. Now, if we compare the two shaded elements above with the pitch tests used in the Seashore and Wing batteries, some interesting comparisons can be made which might throw light on why Memory and Pitch seem to be associated in the present study. The main point is that the situation represented in the shaded portion of the above diagram is also a representation of the situation which obtains in tests of pitch discrimination. In the Seashore Pitch test, a subject is presented with a pure tone of 500 c, p, s.; this is followed by a tone which the subject has to compare with the first one, indicating whether it is higher or lower than the reference tone. In the Wing Test a similar situation obtains. Basically, two consecutive tones are sounded, and the sbuject has to indicate 'up', 'down', or 'same'. Admittedly, in the pitch tests a judgement as to which direction the change takes is required; this is not asked for in the memory tests. However, the comparison of two stimulus tones, and the initial detection of any change, is common to both the pitch and the memory tests. There are two minor differences in the procedures. Firstly, Seashore called for much finer levels of discrimination, up to one two-hundredth of a tone, whereas Wing asked for judgements on the tempered scale. (This difference in the level of discrimination called for affects the present argument in no way whatsoever). Secondly, Wing's compared tones were in fact middle notes in a series of triads. The effects of the differences will be as follows. The finer levels of discrimination called for by any particular item will make that item more difficult than items calling for less fine degrees of discrimination. We might thus expect a person scoring low on the Wing test to score at least as low (if scores were comparable) on Seashore's test. On the other hand the Wing test is rendered more difficult since the tones to be compared are presented along with other tones. We can call these extra tones 'masking' tones,; the effect of such masking tones is to make discrimination tasks more difficult. This situation can also be viewed as a signal detection problem, with the relationship of the stimulus tones to the masking tones described in terms of a signal to noise ratio. (T. Dean Clack, 1967, Ritsma, 1966, B. L. Cardozo, 1967). The analogy between the types of test used in musical test construction and signal detection is not just a fortuitous circumstance: signal detection really can provide a body of experimental data that can help to provide the body of "experimental demonstration" that gives common ground for future test construction. It could be argued that constructors of tests of musical ability have been to some extent intuitively constructing signal-detection tasks.

We have already stressed the importance of 'recognition' in McLeish's study. On the topic of recognition, D. M. Green and J. A. Swets write, (1966) "In order to obtain a more sensitive measure of retention than is provided by a direct measure of unaided recall, a variant of the recognition procedure is often used. In one form of the recognition procedure, the subject's ability to remember stimuli to which he has been exposed is tested by presenting those stimuli intermixed with other stimuli to which he has not been exposed, and by requiring him to state for each stimulus whether it is "old" or "new". (Woodworth and Schlosberg, 1954) The subject, in effect, responds "yes" or "no" depending on whether or not he thinks the stimulus is an old one". In fact, this describes the Memory Tests of Seashore and Wing very accurately, and also test material by many other writers, including Bentley, Drake, Whistler and Thorpe, and others. In fact, if we look at all the items from the tests used in McLeish's study, we can describe them all in terms of signal detection tasks; though some are more amenable to this than others. Signal-detection theory was developed initially as a method for measuring psycho-physical phenomena; however its applications are now much wider and a considerable amount of work has been done using signal-detection as a measure of retention.* (Egan, 1958. Norman and Wickelgren, 1965. Murdock, 1965. Pollack, Galanter and Norman, 1964). Since

* Much emphasis has been given to the psycho-physical nature of Seashore's tests, as contrasted with the approach of workers like Wing. In a signal detection situation,

we can describe virtually all the different sub-tests of musical ability in terms of signal-detection tasks (in the McLeish study, only Wing's chord analysis test, of the tonal items, is difficult to fit into this scheme) it is hypothesised here that all the test items are in fact concerned with the measurement of some aspect of short-term memory, and that the common factor found by McLeish is in fact a memory factor. This hypothesis is advanced on the basis of the observed correspondence between types of test material used in tests of musical ability and situations obtaining in signal-detection type experiments. It is hoped that it will be an aid to understanding better the results of certain selected factorial studies of music tests.

In interpreting McLeish's findings in detail, the above hypothesis can be of help in the following manner. If we look at the tests of Wing and Seashore, we can perhaps observe how the highest loadings are obtained on those tests which we might expect to have the highest saturations of 'signal-detection', and hence of memory. Highest loadings are obtained on the 'Memory' parts of both tests. These items, individually,

* Continued from overleaf.

recognition seems to behave very much in the manner of a psycho-physical variable, however. Green and Swets state, "This memory task is much like the yes-no task of detection in psycho-physics and, understandably, models of the memory process have developed in parallel with models of the detection process. Specifically, recognition memory has long been viewed as a threshold process,....."

have the longest duration of all the 'aptitude' tests, and length is varied whilst 'signal' stays constant, which affects the difficulty of items. By altering length, the signal-noise ratio was intuitively being manipulated. The fact that this type of procedure was clearly a measure of memory is reflected in the names given to tests of this type by test constructors. The similarity between the Pitch tests, which tend to have almost equally high factor loadings, and the Memory tests, has already been pointed out, (p. 21). In the case of Wing's Pitch Test the 'noise' takes the form of masking tones (that is, the interference occurs simultaneously with the signal in each presentation, rather than being spaced, in a temporal sense, around the signal); Seashore's on the other hand is not masked, but calls for finer degrees of discrimination thus making decision more difficult. It is important to note that only the first version of the Seashore test contained any test of harmony or 'consonance'. In the revised version this was replaced by a test of timbre. McLeish's findings with regard to harmony or consonance are thus applicable only to the unrevised version of the Seashore tests. The only other tests in the two batteries which are in any way comparable are Seashore's test of consonance and the Wing test of Chord Analysis, though there are considerable differences here. Of the two tests, the Wing loads the more highly on the common factor; so we would expect, in view of the present hypothesis, that the Wing test would have more of the characteristics of a task in recognition memory than the Seashore. It is fairly easy to see that this is so. The Seashore test of consonance involves the subject in listening to two consecutively played pairs of notes, (i.e. a chord of two notes is played; then after a pause a further chord of two notes is played). The subject must say which pair 'blends better'. The point here is that 'better' or 'worse' is a value judgement; and in an aesthetic sense it cannot be described as a threshold process. The criterion for making the judgement is external to both presentations of the stimulus. In the Wing test, the subject must listen to the presented chord, and then by a process involving tonal imagery, compare a copy of the chord as heard with a series of hypotheses about the notes it might contain. The amount of sub-vocal or even vocal singing of different elements of the chord during the performance of this task suggests that a process of hypothesis and acceptance/rejection is taking place. (It is possible that the subject stores the chord in short term memory; then produces a series of tonal images asking the question, "Is this note in the chord you just heard?"). The criterion for answering is internal in this case, and we can regard the ability to perform the task as a threshold process, in the manner of the tests of Memory and of Pitch.

The remaining tests in each battery are entirely different in conception and purpose. The intensity, time and rhythm tests of Seashore are all concerned with aspects of rhythmic ability, which may be different from abilities measured with tonal material. Even protagonists of the group factor theory have conceded that rhythmic abilities may be different from tonal abilities. Shuter writes, "A separate factor of rhythm seems to lurk in the shadows of several studies where tests of pitch and memory have been highlighted. Both Karlin and Wing acknowledge the possibility of its existence. Inspection of many of the correlation studies shows that the tests of rhythm seem to be the odd man out (McLeish, 1950; Bentley, 1955; Rainbow, 1965). In both McLeish's and Wing's factor analyses, the rhythm tests had the lowest loading in the general musical ability factor. In Franklin's study, the two tests that he adapted from Revesz form a separate factor of their own, unrelated to the Wing rhythm test". The remaining tests in the Wing battery (rhythmic accent, harmony, intensity and phrasing) are in no way comparable with the tests of rhythmic ability in the Seashore battery. The former are thoroughgoing tests of appreciation, which Seashore's are patently not. Examination of the results obtained by McLeish shows that factor loadings for the tests described in this paragraph are somewhat lower than for the three types of test material which were comparable, (i.e. pitch, memory, consonance). It is difficult to account for the fact that the appreciation tests of Wing seem to load more highly on the first factor than do the tests of Seashore. However, if we take the view that rhythmic ability is a separate entity from the main musical factor we can account for the low loadings obtained by Seashore's rhythm tests. The higher loadings on the Wing appreciation tests are more difficult to explain; but we may note that Franklin's study showed that the Wing rhythm test did not measure the same thing as his own adapted tests which formed a separate factor of their own. It seems likely therefore that the Wing rhythm test does not attempt to measure rhythm in the same way that Seashore's does. It has already been suggested that the common factor in musical ability tests is an aspect of short term recognition memory. From the factor loadings on the Seashore test and Wing test found in McLeish's study, we might also observe that as a general rule, items which do not involve tonal material tend to have lower factor loadings than those which do. Thus, though Wing's appreciation tests do not load as highly as tests of memory or pitch, they still have higher loadings than Seashore's rhythmic measures, which involve no tonal material. If we regard the common factor as being a memory factor for material of a tonal nature, then although Wing's last four tests involve no internal criterion for discrimination, we might expect slightly higher loadings on these since they involve two compared presentations of tonal material, whereas Seashore's do not involve tonal material.

Conclusions from the Factorial Studies.

The main conclusions from factorial studies of musical ability are not incompatible with the conclusions reached by McLeish. It has been shown that a common factor is present, even in tests as atomistically orientated as Seashore's battery, and this factor is apparently identifiable as the same one present in the Wing tests. However, it is perhaps unjust to say that Seashore was wrong on the grounds that his tests were not mutually exclusive; a considerable degree of specificity exists within the tests, and there is evidence to suggest that in fact tonal and rhythmic abilities are more or less specific. Neither is Wing entirely vindicated in his comment that the Seashore tests were of a "doubtful nature,....

for they did not involve appreciation in any marked degree", since his own appreciation tests were shown to have loadings on the common factor which were without exception lower than the loadings on those of his tests which did <u>not</u> involve any appreciation.

As far as test batteries, and the degree to which different items correlate with the common factor, are concerned, McLeish's conclusions as to what are the principal 'components' of 'musical cognition' are perhaps not entirely satisfactory. Certainly, in the structure of musical cognition, tests of pitch and of memory seem to give the highest loadings; but it has been suggested here that in fact tests of pitch and of so-called memory are in fact both tests of memory for tonal materials, and both can be viewed as tasks in pitch discrimination. The conclusion that chord analysis is also a major component cannot be accepted without comment. Seashore's tests <u>did not contain any items involving chord analysis;</u> in the first version of his tests, subjects had to choose the pair which 'blended better', this being a response to the sound created by a fusion of tones, which can in no way be viewed as an analytic process. (No test of consonance or chord-analysis appeared in the revised version) Also, when the tests were analysed together, Wing's chord analysis test correlated with the group factor no more highly than certain of his appreciation tests.

Musical Ability - A Mental Process.

In view of the assertion by Green and Swets that recognition memory can be treated, as far as its measurement is concerned, as a threshold process, the argument against Seashore that his tests were 'psychophysical' in nature, with the implicit assumption that a psychophysical measure was inappropriate in this situation, needs re-examination. (An excellent example of the application of psychophysical techniques to musical problems is given in Trotter, J. R., 1967). Firstly, insofar as Wing's tests are tests of recognition memory, they are also psychophysical tests. (This clearly does not apply to his appreciation tests, which involve external criteria) It is necessary therefore to redefine this particular attack on Seashore by saying that, in testing the merely peripheral functions such as sensory thresholds he chose the wrong thresholds, this accounting to some extent for the rather low validity of certain of his tests. Further evidence for this is provided by the fact that Seashore's Tonal Memory test, which is the one test that does in fact measure by psychophysical means a central process of memory (as distinct from a sensory threshold), has far greater validity than any other of his sub-tests. Psychophysical methods then are not innappropriate in themselves. The main point must be that musical ability is a mental ability and not a peripheral capacity, and that whatever method of measurement we choose, it must measure something more than simple sensory thresholds if high validity is to be obtained. The fact that musical ability is a mental and not a peripheral function is crucial from the point of view of test construction, and Seashore's pre-occupation with the 'nature of the medium' (that is, the physical properties of sound) led him to an approach which concentrated on ways in which the physical properties of music were detected by people. As a result the bulk of his tests measure sensitivity to <u>any sound stimulus</u>, rather than response to specifically musical sounds as distinct from noise. Only his test of melodic memory, of the tonal items, is specific to musical material. Mursell stressed the essentially 'mental' nature of musical ability, and wrote as follows. "This crucial fact, that we hear mentally created patterns rather than imposed sensations - that the mind selects and organizes and gives shape to what we hear - is the foundation of all musical organization and the secret of the expressive possibilities of music". (Mursell, J. L., 1937)

Further to this, Madison writes, "Mursell, therefore, sees a gap between responses involving simple sensation and those which perceive musical relationships. This gap, he points out, can be accounted for only by mental activity which functions by means of selection and synthesis of sound stimuli in the conception of meaningful musical configurations. Throughout his whole work Mursell sees great danger in confusing the laws of the physics of sound with the psychological laws governing the perception of musical values expressed through the medium of sound stimuli". And later, "A certain school of psychological thought adheres to the assumption that, by virtue of the medium of musical expression, musical talent is controlled and made possible through sensitivity to physical differences of the sound wave. Another school of psychological thought regards the true basis of musical talent as being the power of mental synthesis of the materials and structure of music as expressed in this medium of sound. This school does not discount the importance and need for sensory keenness".

An Attempt at Definition.

Much of the preceding argument has been an attempt to find some sort of rationale behind studies of musical ability tests, and particularly factorial studies. The hypothesis has been put forward that the common factor found in many studies is some aspect of short term memory. Initially, this hypothesis suggested to the writer that in fact the short term memory factor was general; with the implication that the common factor referred to by writers as the 'musical factor', or 'musical cognition', or something similar, was in fact nothing to do with music but was a general rather than a group factor. This conclusion was later rejected when experimental evidence was obtained showing that ability to remember certain musical materials did not correlate with ability to remember other non-musical materials. A common factor for musical ability, as distinct from other abilities seems a likely explanation. It seems probable, therefore, that tests of musical ability have in common a short term memory factor which is specific to musical materials, and the following attempt at definition is put forward:-

> As measured by tests of musical ability, the common musical factor is a short term memory factor which is specific to musical materials. It enables persons to hold in short term memory certain musical elements, in the form of musical images, long enough to permit a process of comparison and recognition of subsequent elements to take place.

It is possible that a separate short term memory factor exists which is specific to rhythmic, as distinct from tonal, materials.

Conclusions from Section 1.

Accepting the above definition as a hypothesis, it remains to propose ways of devising test material that will help to assess its validity. Very little of the existing test material can help to throw further light on the matter, since the majority of it is far too factorially complex. J. P. Guilford, 1952, writes, "Too many experimental variables are factorially complex. Rotations and interpretations would be much simplified if each variable were of complexity one; that is, if it measured only one common factor to any appreciable extent. This is an ideal that we achieve in test construction only once in many attempts. There is little excuse for taking almost any variable that is handy. Such variables, where there has been no effort to restrict them, are very likely to measure two or more common factors".

We have above a definition of musical ability in terms of short term memory for specific materials. It would be advantageous, therefore, if by manipulating short term memory variables we could produce a test battery which, in view of Guilford's comment, produced substantially only one common factor. We might then suggest that the common factor was related in some way to the short term memory variable. It would also be most important, in the present context, to show that such a test battery met certain criteria as regards reliability and validity. If such a battery gave good reliability and validity we might be tempted to hypothesise something about the nature of musical ability in general, as distinct from musical ability as measured in the specific test situation.

It has been noted that in the study by John McLeish, moderate loadings on the group factor were obtained on Wing's tests of appreciation. The same factor also appears in the Oregon Music Discrimination Test, which is essentially a test of musical taste and appreciation. The point should be made that in such cases high factor loadings might give a distorted impression of the value of certain measures. In the case of the Oregon Test, McLeish's results show that this test has the highest loadings on the common factor. McLeish concludes, that "the Oregon test is the best measure of musical ability, insofar as it is most highly saturated with the factor of musical cognition. In fact, the particular musical talent measured by the Wing and Seashore tests is better measured by the Oregon test, even though this is a piano test which uses classical pieces". The Oregon test involves the subject in listening to extracts from the compositions of 'accepted composers'. Two renditions, one of which is mutilated, are played, (an earlier version in which no less than four extracts were heard was discontinued as being too difficult) and the subject must indicate which is the original. Clearly, short term memory is required in order to compare the first rendition with the second. However, the criterion for judging 'original' or 'mutilated' is external to both presentations, and influenced by environment and experience. Thus, the factor of short term memory for musical material, whilst essential to the task, is not in itself sufficient to permit the performance of the task. Hevner's finding that her tests discriminated well between psychology students and advanced music students is hardly surprising; and McLeish's high factor loadings do not imply that this kind of test material is the best for incorporation in a test battery. Of the Oregon Tests, Shuter writes, "The Oregon Tests have usually been regarded as tests of taste and appreciation, as distinguished from ear acuity tests. However, ability to perceive the differences between the accepted and distorted version is obviously required. Moreover, building up a listening repertoire of good music with which to compare the versions must partly depend on general auditory efficiency". In this context, Shuter clearly meant to imply that these tests were to some extent acuity measures. The main point however centres around the need for "building up a listening repertoire of good music with which to compare the versions". Without such a repertoire, the ability to discriminate between the versions is no help at all in performing the task. This clearly places severe limitations on the circumstances in which such a test can be used. In short, the Oregon tests include a strong experiential/environmental factor which is not linearly independent of the musical factor.

In devising any test material for use in picking out potential musical ability (musical aptitude), items between which the intercorrelations are the result of combinations of factors such as appreciation and experience/environment are best avoided. Guilford writes, "There are a number of situations in psychological investigations in which specific and error variances actually contribute to intercorrelations where they should not be allowed to do so". In selecting items for any test battery it is best therefore not to be too easily seduced by high factor loadings. The proof of the pudding lies not in the recipe, but in the eating; and the proof of a test battery lies not in factor analysis but in reliability and validity.

Several definitions of 'musical ability' have been advanced by different writers. Some of these are of little help in suggesting ways in which 'musical ability' might be measured or quantified. We can define musical ability as something which tests of musical ability measure, however. It then remains to examine certain factorial studies in order to discover what it is that musical ability tests measure. Unfortunately, many such studies bring to light various factors whose nature and definition is no more precise than the original definitions of musical ability.

Several studies, however, have produced high loadings on tests of memory; and a few writers have chosen to identify common factors as being some aspect of memory. Also, the situations obtaining in by far the greater proportion of test batteries are exact parallels of situations obtaining in signal detection experiments; particularly those concerned with the measurement of recognition memory, which is seen as a threshold process. The hypothesis that musical ability, as measured in tests, is in fact a type of short term recognition memory specific to musical or quasi-musical materials, is put forward. It remains, therefore, to see if, by manipulating test material in accordance with the above hypothesis, we can produce a pilot battery that satisfies certain criteria of reliability and validity.

SECTION 2.

The Form of the Present Test, and its evolution.

The Proposed Battery: An Outline.

The battery of tests proposed here differs in its fundamentals from existing types of psychologically orientated tests, and also from the types of test material used by musicians and music teachers in their selection procedures. The present battery consists of four sub-tests; Part 1 is styled melodic memory. The task, as it was first envisaged, involved listening to a short piece of music (statistically derived) consisting of perhaps four or five notes, and trying to hold these in memory. The short tunes were to be followed by a rather longer piece of tune containing twice as many notes. The task involved locating the short tune in the longer one. In general terms it was hoped to make subjects report (a) whether the short tune occurred in the longer one, answering "yes" or "no", and (b) make some kind of judgement about where the short tune was located within the longer one, whenever the answer to part (a) was "yes". It was thought that a task based along these broad lines would involve the recognition of the overall shape of short tunes, without placing undue emphasis on any single tone in the material used.

Part 2 of the test battery is called Pitch Recognition. This task, in contrast to all existing tests of pitch, demands the locating of a previously heard tone in one of several different sweep frequency tones. Only one of the sweep frequency or 'glide' tones would pass through a frequency the same as the tone heard at the start. The emphasis here is placed on localising a given tone within a broad band width; this contrasting sharply with existing pitch tests which place a premium on finer types of pitch response, but do not demand 'localisation' of the same type.

Part Three of the test battery is an interval measure, titled Auditory Transposition. As originally conceived, subjects would be presented with two temporally separated tones of different frequency. They would then be presented with several more pairs of tones, and asked to pick out which of these pairs had the same tonal separation as the pair first heard. In order to obtain a correct answer, subjects would have to respond not to the absolute frequency of the tones used, but to the frequency ratio of the two tones used. This ratio can stay constant, and be recognisable as a constant interval, even though the absolute frequencies of the tones in the pairs compared may be very different.

The fourth part of the test is a Rhythm Test. This is an attempt to embody in a single sub-test a complex task which can give a good indication of rhythmic ability. Basically the task involves making a match between a steady metre in 2/4, 3/4, 4/4, or 5/5, and a more or less complex, subdivided rhythm. This type of task is a very important one in musical performance, which often demands the reproduction of complex subdivided rhythmic patterns overlaying a stated or implied steady metre. Patterns of three over four, or two over three are fairly common examples of this.

In brief outline, the above formed the basis for the test battery, though certain of the tasks changed in detail as a result of preliminary test runs. It should be noted that there is no test of consonance (harmony) included, nor any test that makes use of consonant material. Also, the tests used are in no way intended to be a total coverage of all aspects of musical ability; in particular, no account is taken of motivational factors, which are considered beyond the scope of this battery. Finally, the rhythm test is not an exhaustive means of assessing rhythmic abilities; rhythmic patterns can be described in terms of amplitude or accent, temporal spacing of elements, and duration of elements. The present test incorporates only the first two of these;

the omission of duration of elements is not serious because, provided temporal spacing and accent are kept constant, the basic underlying metre and the imposed complex rhythmic pattern remain essentially unchanged even when duration of individual elements is altered. On the positive side there are advantages to using punctate stimuli, not the least of which is clarity.

It remains now to describe in some detail the evolution of these four types of sub-test, and the reasons why tests of this type were chosen in preference to others.

Types of Material Available.

The first fully standardised tests of musical ability produced in this country were the tests of Musical Intelligence devised by Herbert Wing. In assembling his battery he chose tests of Chord Analysis (how many notes in a chord), Pitch Change (detecting change 'up' or 'down' of one of the tones in two consecutive chords, or no change), Memory (detecting an alteration of note in a short melody). In addition to these three tests he also used four more tests of a rather different nature, in which subjects had to make value judgements of one type or another. These were Rhythmic Accent (choosing the better rhythmic accent in two performances), Harmony (judging the more appropriate of two harmonizations), Intensity (judging the more appropriate mode of varying loudness-crescendo, decrescendo, etc. - in two performances of the same melody), and Phrasing (judging the more appropriate phrasing - grouping of notes by pauses, legato and staccato playing, etc, - in two performances). In arriving at this particular selection of test material, Wing had reason to reject certain other methods of assessment. His reasons for rejection are of interest, since in some cases they apply almost directly to the present study.

Wing gives the following comprehensive list of the types of tests set by musicians, divided into three categories which he calls Aural Tests, Paper Work, and Performance. (B. J. P. Mon. Supp. XXV II, 1948.)

(a) Aural Tests:

1. Intervals.
2. Chord Analysis.
3. Cadences.
4. Discord Resolution.
5. Key Changes.
6. Time Pattern Dictation.
7. Dictation using tones:
   (i) Melodic Pattern Dictation.
   (ii) Harmonic Pattern Dictation.
8. Recognising Music.
9. Memory.

(b) Paper Work:

10. General Musical Knowledge.
*12. Rhythm.
13. Melodic Shape.
14. Harmony.
15. Fitness.
16. Creative Ability.

* In the reference quoted above, no item bearing the number 11 appears.

(c) Performance:
17. Intensity.
18. Phrasing.
19. Pace (accuracy and variation).
20. Emphasis of a Part - implying an appreciation of form.
21. Pitch Accuracy - for variable pitch instruments. (a test of ability rather than appreciation).

It should be noted initially, that all items in parts (a) Aural Tests and (c) Performance are in fact applied by musicians as individual rather than group tests. The success which is achieved with such methods is perhaps due as much to this fact as to the efficiency of the tests themselves.

Wing devised a battery of tests, dealing with almost all the above types of material and concluded that not all these 'musicians' tests could be adapted profitably to a more psychologically oriented test procedure. Some of his reasons for rejecting certain types of item are sound, and as valid today as when they were first written. Clearly certain of the above tests rely on either past musical experience or acquired knowledge to an extent that makes use in a so-called culture free test impossible. In the first place one can discount all tests demanding answers in the form of musical code, such as note names, time values, or any aspect of musical notation. Any person without training in these skills, or knowledge of the terms used cannot hope to do well in the tests. Even if preliminary instruction in these types of things is given the situation is still unsatisfactory since differences between scores could still be accounted for by the differing degree to which subjects had mastered the material. This argument prohibits the use of any type of musical dictation test, items 6 and 7 in Wing's list; and also item 5 except under very modified circumstances, since the concept of 'key' is a specifically musical one and we cannot simply ask our subjects to indicate when 'a key change takes place' or similar. Care must also be taken that any items chosen do not favour certain individuals over others in respect of specifically musical experience, or of more general home background which in itself can lead to differences in musical contact or experience. Much care must be exercised in satisfying this requirement as influences here are much more insidious, and almost all types of material will be influenced at least to some degree by cultural/environmental influences. Items 3, 4, 5, 8, 10, 15, 18 and 20 are all likely to be influenced in marked degree by these factors, even if tests are modified to minimise such effects.

Item 3, cadences, for example is particularly liable to be affected by learning, especially when subjects are asked to make judgements about 'completeness' or 'incompleteness'. The Lowery cadence and phrase tests have been mentioned in an an earlier chapter. Of these tests Shuter says; "Cadence tests are difficult to apply to subjects without musical training owing to the difficulty of describing them and because, in any case, two chord cadences present a certain ambiguity of key'. Wing also found these tests unsatisfactory for this reason and also the fact that in his early battery he found them too easy; these two facts are admittedly a little difficult to reconcile.

Item 4, discord resolution, is also unsatisfactory. Early theories of hearing implied that parts of the inner ear acted simply as natural resonators. 'The ear is a tiny piano' wrote Mrs. Spencer Curwen.* Helmholtz also clearly saw the basilar membrane as an elastic body which operated after the fashion of one of his resonators;.....'the sonorous vibrations of the air in the outer auditory passage are finally transferred to the membranes of the labyrinth....' (Sensations of Tone, Chapt. VI, Part I) These early theories led, naturally enough, to certain views on the subject of consonance and dissonance; Hemholtz for instance writes (pp 330)

* Mrs. J. Spencer Curwen; Psychology Applied to Music Teaching. London: J. Curwen & Sons Ltd. (No date)

"If the eye is a little camera, the ear is a tiny piano, a piano with a keyboard for the air to play upon, with 11,000 strings behind the keyboard, and with a damper to stop the movement of the strings after they have sounded". pp 21, Mind and Body.

'When voices move forward melodically in part music, the general rule is that they must form consonances with each other. For it is only as long as they are consonant, that there is an uninterrupted fusion of the corresponding auditory sensations. As soon as they are dissonant the individual parts mutually disturb each other, and each is a hindrance to the free motion of the other. To this esthetic reason must be added the purely physical consideration, that consonances cause an agreeable kind of gentle and uniform excitement to the ear which is distinguished by its greater variety from that produced by a single compound tone, whereas the sensation caused by intermittent dissonances is distressing and exhausting'. One of the more interesting things about this extract is the implicit assumption that sounds can in fact be readily classified, even with complete disregard to any context, as 'consonant' or 'dissonant'. Modern views on the function of the cochlear and the basilar membrane suggest that the above are an over-simplification, (von Bekesy, 1960; Whitfield, 1967). Two pieces of experimental evidence may provide food for thought here:- J. H. Dewson (unpublished) used monkeys in an operant situation in which they could hear two continuous pure tones. They were taught to move a lever by which means they could vary the interval between the tones, a task in which they acquired a certain skill. Attempts were made to see if the monkeys preferred any intervals over others. In fact no preferences were shown, and no evidence could be obtained to suggest that they found so-called 'dissonances' any more exhausting or distressing than so-called 'consonant' intervals. The suggestion here is perhaps that the perception of consonance and dissonance is based on something other than natural processes at a cochlear level as suggested by Helmholtz. Also, work by A. Hickman (unpublished Ph.d. thesis, Manchester) points to a similar conclusion. In one of his experiments, subjects were played various chords and asked to indicate how many separate tones made up each chord. However, the chords were compounded of mixed pure (sine-wave) and complex tones. The complex tones were synthesised from phase linked pure tones, at frequencies identical to the expected natural occurrence of fundamental and upper partials. Subjects reported more tones to be present when the complex tones were used than when only sine wave tones were present. The important point is that he found his results unaffected by the synthesising of tones in which the artificial upper pertials did not correspond to the natural harmonics of the fundamental. If any part of the auditory system operated, as far as consonance/dissonance is concerned, in the fashion of a free resonating body, then more tones should be heard for the tones with the 'false harmonics' than for tones compounded from 'true' harmonics. This did not take place.

In the absence of any convincing evidence that consonance/ dissonance stems from within the system, we must assume that it comes from outside the system. In this case, experiential and learned factors will have an influence. There is musical evidence to suggest that the concept of consonance is very much influenced by cultural factors, and the climate of composition at various epochs. The use of natural and flatted sevenths is common nowadays; also flatted fifths (f sharp in a 'c major' chord for instance) can be regarded as pleasant harmonic variations. There are a host of other chord positions that illustrate this, and even whole chords made up entirely of tones with competing upper partials are acceptable as true and good examples of harmony. Certainly they are not the distressing, exhausting experiences that Helmholtz suggests. In summary, then, the step from consonance to dissonance is not the simple one it might first appear, and can not be adequately explained in terms of beats between upper partials. Experience of, and exposure to, developed types of harmony leads to increased 'tolerance' for the new forms, and what was dissonant yesterday is not so today. Thus any test depending on an assumed clear-cut distinction between consonance and dissonance (concord and discord) is unlikely to prove satisfactory, and will be inherantly culture biased.

Item 5, Key Changes, is unsatisfactory since the concept of key is a purely musical one and non-musicians can be expected to have only a very poor notion of what key is. We cannot therefore ask them to indicate when key changes take place. As already stated preliminary instruction in the musical terminology is not desirable. Also the concept of key, bound up intimately with tonality and shifts about tonal centres is not one that can be taught in five minutes prior to test administration. Briefly then, many subjects would certainly be unclear as to what precisely they were supposed to be looking for. On tests depending upon detection of key change, Wing writes, 'The test on key change, ...gave results which showed a satisfactory scatter and which indicated that it had promise of being a good diagnostic test. However, it appeared essential to give a short introductory lesson which included practice with suitable examples, before applying the test, in order to make clear to the listeners exactly what was required of them. This considerably lessened its value as a psychological test, and it was therefore rejected'. (pp 48).

The implications of tests along the lines of Number 8, Recognising Music, are surely very clear. This type of test can not be viewed in any way as one of ability or aptitude; but only as a test or measure of past experience. Wing makes the point that tests of this type 'were found to be particularly subject to the effect of opportunity'. As further evidence he points out that little agreement was found between measures of this type and scores on more satisfactory test material. He concludes that they were 'probably inefficient'. (pp 44) The effects of home environment upon musical opportunity have already been mentioned.

Much of what has been said about Test No. 8 can be applied to Test 10, General Musical Knowledge. Though in some situations, tests of this type may have undoubted value (as in assessing previous musical experience) in the present context they have no place.

The idea of fitness (Item No. 15) is closely bound up with much that has been said already. The making of judgements about the fitness or appropriateness of certain aspects of a performance implies that the person making the judgement has a background knowledge, based on previous experience, against which he can make his judgement. Without such a background, the situation is meaningless. Any value judgement, if it is to be useful, demands such a store of experiences. Basically, a test of this nature merely measures the correspondence between the subjective views of the testee and the subjective views of informed judges. The difficulty here is that the musical basis upon which the experts make their judgements is very highly culture specific. What is 'fit' in European music is not necessarily fit in the music of other cultures. For example, the Western concept of key, or of cadence, cannot be generalised in any way to Indian music. The whole concept of tonality in Indian music is different, such that Western listeners hear the music from an entirely different tonal centre than the Indian listener, (i.e. it is played in a key different to the one we think it is in). The intervals in the scale (mode) are also different from the Western system based on the tempered scale. These are, however, extreme differences between the two culture patterns, and it might be argued that such tests could nevertheless work quite adequately within a single culture, or between cultures with some moderate degrees of difference. This also is unlikely to be the case since, even within a single culture, there are sub-cultural differences that render this type of item still unworkable. Shuter (1964) related the Wing scores of 189 junior musicians from the Royal Marines School of Music to the socio/economic level of the father's occupations. Musicians whose fathers were rated as being in the highest social class gained over twice as many scores in the two top classes of the Wing Tests as those whose fathers came from the lowest social class. The findings on this point are by no means unequivocal, but Shuter (1949) sums up the situation as follows:- '(Rainbow) found that the correlation between socio-economic status and home environment was about .3. This would confirm the everyday observation that there is a tendency for musical activity in the home to be related to socio-economic status but that higher social status and a regard for music do not always go together'. Generally we can safely say that children of parents in higher social orders stand a better chance of being exposed to more satisfying forms of music than children of parents who are manual workers or similar; perhaps through encouragement to learn an instrument, devoting funds to the purchase of suitable instruments, visits to theatres or concerts, or participation in local groups of various types.

In addition, different people prefer, and are subject to, many different types of music within our culture. Some may prefer to listen to classical music, others to 'pop' music, other to 'folk' and others to nothing in particular. This creates great difficulties in the devising of 'fitness' tests which are equally appropriate to all forms. 'I do not like classical music because it is boring and miserable'; most music teachers must have heard something similar to that said by their young charges from time to time. Faced with such a remark we tend to put down the lack of partiality to what we regard as superior forms to lack of sufficient and proper exposure to, and experience of, such forms. Note, however, that Herbert Wing writes; 'Jazz music was not included, as this would be unlikely to yield examples of really good harmony,........and would waste the children's time if they were listening to poor music'. Perhaps we can put this attitude down to similar reasons. Certainly the choice of one type of music to the virtual exclusion of all others cannot be justified in these terms. In this instance, the author's sense of 'fitness' does not extend beyond the field of classical music, and in judging jazz he suffers from exactly the problem described above; certainly the generalisation that jazz contains little, if anything, of harmonic interest and is generally 'poor' cannot be substantiated. The reasons for not using tests of this type then are twofold. Firstly, there are within any society different sub-cultural groups who can be to some extent differentiated in terms of musical experience; and secondly, even within sub-cultural groups different people have different preferences, and opportunities. Thus, without rigorous selection procedures we cannot ensure that any sample of people have a common store of musical experiences on the basis of which to make value judgements concerning fitness or appropriateness.

Item 18, Phrasing, cannot be used for many of the reasons given above. That is, good or bad phrasing, or correct or incorrect phrasing, is under the influence of subjective opinion, based on what is normal or habitual within any culture. Also, when applied as a group test the situation becomes almost the same as in items involving fitness. Certainly, anyone not familiar with the type of material used in existing tests of this type is at a distinct disadvantage. Note that Seashore has described the use of glissando by negro singers. This is used as an aesthetic ornament; many Western musicians would tend to regard such slurring of sung notes as slovenly or in bad taste. Similarly, the conception of what is acceptable, or even desirable, in traditional jazz, or some other type of music, might clash sharply with a classical musician's conception.

Item 20, Emphasis of a Part, is described as implying an appreciation of form. Tests involving appreciation, or discrimination between good and bad form, demand previous experience of analogous situations, or a store of musical experiences sufficiently large to allow the formation of concepts of good or bad form. Also, the testing methods advocated by the present writer are in no way intended to be measures of appreciation, so items of this type cannot be included.

In the light of the above arguments, only ten types of item emerge from Wing's original list of 21 as being suitable for development in a psychological test battery. These are items 1 (intervals), 2 (chord analysis), 9 (memory), 12 (rhythm), 13 (melodic shape), 14 (harmony), 16 (creative ability), 17 (intensity), 19 (pace), and 21 (pitch accuracy). There is a fair correspondence between this short list and the list from which Wing drew his final test battery. Note, however, that no test of phrasing is present in the list, though tests of phrasing were used by Wing, and also Lowery (1929); and that item 1, intervals, is retained even though this test was rejected at a late stage by Wing. It is useful to remember that in deriving his own short list, Wing did a comprehensive testing of all types of musical tests, and produced a quantity of statistical data on the basis of which much of his rejection procedures were carried out. Altogether, he used twenty-six different tests in his preliminary survey, and in so doing provided an invaluable basis for other researchers to work on.

Selection of Items for Use in the Present Battery.

The initial list of 21 types of testing material has been reduced to ten; mainly on cultural, environmental or experiential grounds. Statistical evidence on this is provided by Herbert Wing, and more detailed examples of this will be given later. However, the restrictions imposed on group tests in terms of length and time available make even a battery of ten items too long to be workable in the present context. It was thought desirable therefore to reduce further the number of items since (a) a test utilising all the ten remaining items would be too long, and (b) there are reasons to suppose that not all the ten items are completely satisfactory as potential test material. The reduction in the number of items took place in the following manner:-

Briefly, reduction took place on the basis of three sources of information:

1. Types of item contained in existing tests and measures, and information on the derivation of such items.

2. The literature on the psychology of music, and the more specific literature on the function of the ear and experiments in hearing.

3. A survey carried out by the present writer on a sample of high grade musicians and music teachers to find out what musical abilities they thought most pertinent to the performance of certain specific musical tasks.

On the basis of the above, it was thought desirable to eliminate certain of the remaining items, and to bring together certain others which had a degree of redundancy; leaving only items concerned with memory, pitch, interval and rhythm.

The Case Against Harmony and Chord Analysis.

The most striking omission from the present battery when compared with other batteries is items involving 'harmony', or any material using simultaneously produced sounds. Reasons for eschewing anything involving judgements about 'consonance' or 'dissonance' have already been given, but there are important objections to any material using simultaneous sounds even when subjective judgements about consonance or dissonance are not called for.

The proposed battery of tests is intended primarily for use in junior schools, that is with children in the age range of approximately 7 - 11 years. Bearing this particular age group in mind certain evidence concerning the use of harmonic material (material using simultaneously produced sounds) is of particular interest. Wing's tests of appreciation, concerning rhythm, harmony, intensity and phrasing have a certain musical value and use a certain amount of harmonic material. On these parts of the tests the scores of children below the age of 11 do not exceed chance level, except on the rhythm test where chance level is exceeded at age 10 (Shuter, p 84). This fact clearly stems to great degree from the very nature of tests of appreciation, and cannot be attributed to the use of harmonic material. However, there is the implication here that differences between consonance and dissonance are not apparent to most of the children with whom this study is mainly concerned, since chance scores were not exceeded on the test of harmony.

On this point, Shuter says, 'It is generally held that most young children have no great appreciation of harmony, finding 'every harmonic accompaniment equally good, whether consonant or dissonant'. Valentine, (1962), writes, 'No appreciable preference for concords before discords is discernable before the (average) age of 9....'. In other words, the type of spontaneous musical activity in which children participate is not characterised by its harmonic content. There are several studies of the development of musical abilities with increasing age, all serving to confirm that the music making of young children, whilst containing material of rhythmic interest and melodic interest, is weak in harmony. In the light of these facts it becomes difficult to justify the use of harmonic material in tests for this age group. The fact that Western music, as written and performed by adults, is characterised by its harmonic richness perhaps misleads some writers into thinking that any test battery must contain a test of this, and that its omission makes the battery in some way incomplete. This is not the view taken here. The musical development of the young child is well described by Shuter. Development is broken down into three main stages which she calls 'The Earliest Years' (age below 6), 'The Middle Years of Childhood' (age between 6 and about 10) and 'Music in Adolescence' (age about 11 or 12 up to adulthood). Although these stages are not clearly defined, and cannot be, no mention is made of the development of harmonic skills in 'The Earliest Years', and only in the second chapter in order to point out that no great skill in harmony begins to appear till the age of about 10 in the normal child.

The fact that sense of harmony emerges later than other abilities raises certain other problems concerned with experience; in particular, why the ability should emerge so late. It seems unlikely that an ability of this type, after being absent during the early years, should emerge at a later stage purely on a developmental basis. There is likely to be a considerable cultural influence leading to the emergence of harmonic abilities. Otherwise, how do we explain why the type of harmonic ability to emerge is always of the type present in the society in question. A final point here is that children of different nationalities reared in this country find our own music natural and the type of music present in their native country often strange; this is hardly surprising, if one takes the view that cultural influences play a major role in musical development.

There are cross-cultural reasons for not including any harmonic material. The music of young children bears a resemblance in some ways to the spontaneous music making of primitive peoples. In particular, the use of chants and simple songs, often produced on the spur of the moment and linked closely to some activity, seems a feature of children everywhere. Shuter cites Moorhead and Pond (The Music of Young Children) Chant is 'the most primitive musical art form, for such it is sui generis, to be found among children and, indeed, among men in general. It is part of the living experience of primitive peoples everywhere ... as a primitive, pagan, unsophisticated musical expression arising from those things which the child feels instinctively to demand such expression'.

The main point being made here is that certain aspects of so-called 'unskilled' musical performance are common to many cultures Similarities have been shown between the primitive music of our own society (e.g. the spontaneous chants or songs of young children) and similar music in other societies; even though the formalised and developed musical forms in those societies are very different from our own developed forms. In particular, certain features of melody and rhythm seem to be cross-cultural, both in the societies characterised by primitive music, and in the primitive music of societies with more formally stylised forms. Now, whatever direction later musical development is likely to take as a result of difference in formal musical styles or conventions, these simple tasks form a basis upon which all such development takes place. We have here, then, a strong case for supposing that these are something near to being the bedrock from which develop all types of musical abilities. If we wish to make culture free measures of potential musical development, we must use such basic material. Harmony, and in particular the Western idea of it, just does not qualify here.

A certain amount of evidence exists showing similarities between the music of certain cultures, though the range of cultures from which it is drawn is perhaps not as wide as one would like. In particular, the so-called primitive music of many cultures is remarkable for the lack of emphasis placed on harmony. On the other hand, rhythmic complexity is often far greater than anything found in Western Music.. The music of Islam, India and the Far East all give comparatively little importance to harmony. In Indian music, harmony is supplied only incidentally, by the sympathetic resonance of strings; there is no structured harmony. Similarly, the ancient Greeks used no consonant material; the Greek word 'harmonia' from which our present word is derived, was used simply to imply a succession of tones, rather than simultaneous 'chords' or similar consonant material.

In addition to cross-cultural reasons for not using harmonic material, there are also practical reasons stemming from the physical nature of sound. These raise problems about what kind of answer is really right or wrong in situations like Wing's or Bentley's chord analysis tests. Briefly the problem concerns interpretation of subject's responses in view of the complex sounds generated by chords which ostensibly consist of only two or three notes. Whenever any chord of two or more notes is played, a number of extra tones, or combination tones, are produced. Helmholtz describes these tones as follows:- "The pitch of a combinational tone is generally different from that of either of the two generating tones, or of their harmonic upper partials. In experiments, the combinational are readily distinguished from the upper partial tones, by not being heard when only one generating tone is sounded, and by appearing simultaneously with the second tone. Combinational tones are of two kinds. The first class,....., I have termed differential tones, because their pitch number* is the difference of the pitch numbers of the generating tones. The second class of summational tones, (have) their pitch numbers equal to the sum of the pitch numbers of the generating tones". These tones are not simply subjective phenomena; they really exist and their amplitude can be measured. When the generating tones are produced from a source having a constant output for the duration of each tone, combinational tones can often be heard quite distinctly. Such a source would be an

* The pitch number of a note is commonly called the pitch of the note. By a convenient abbreviation we often write a' 440, meaning the note a' having the pitch number 440; or say that the pitch of a' is 440 vib. that is, 440 double vibrations in a second. The second term, frequency, which I have introduced into the text, as it is much used by

organ or an electronic signal generator. The situation vis a vis combinational tones is less clear where a piano' is used, as the output from a struck string decays. This will be discussed in more detail in a later chapter.

The situation is further complicated because chords containing certain intervals, or played in certain positions on the piano, seem to contain more notes than others, even where this is not the case. The following extract is taken from Teplov, discussing certain findings of Stumpf, who asked children how many notes they heard in certain chords. "Il suffit en effet d'avoir quelque habitude du timbre des sons du piano pour etre en etat de juger, a partir du seul critere de timbre, si l'on entend un son unique ou deux. Les enfants, d'ordinaire, suivent precisement cette voie lorsqu'ls essaient de resoudre le probleme pose par Stumpf. C'est ce que prouvent tres clairement les experiences faites avec des enfants du plus jeune age sous la direction de Stumpf lui-meme. Ces resultats peuvent etre decrire ainsi: l'octave est percue le

Continued from overleaf.
* acousticians, properly represents the number of times that any periodically recurring event happens in one second of time, and, applied to double vibrations, it means the same as pitch number". A. J. Ellis, in The Sensations of Tone.

plus souvent comme un son unique, la quinte comme constituee de deux sons, et en ce qui concerne la tierce et la seconde majeures, les enfants les percoivent en general comme constituee de trois sons, la seconde majeure l'etant meme assez souvent comme faite de quatre. Ceci s'observe aussi bien chez les enfants "tres musiciens" du plus jeune age que chez les enfants "peu musiciens" des ages plus avances. Ces experiences prouvent directement que les enfants repondent a la question: "Combien entendez-vous de sons?" en usant du critere du timbre, et non d'une veritable analyse auditive. L'analyse auditive ne peut aucunement decouvrir dans un accord plus de sons qu'il n'en existe en fait. Or nous voyons que dans les secondes et les tierces majeures les enfants entendent generalement plus de deux sons. Visiblement, leurs jugements n'ont pas pour objet le nombre des sons distingues par l'oreille mais la planitude relative de la sonorite."

Translation:

It is necessary in fact to have some familiarity with the quality (timbre) of the notes of the piano in order to be able to judge, aside from the single criterion of timbre, if one hears one sound or two. Children, ordinarily, use precisely this method when they try to solve the problem set by Stumpf. It is this which is demonstrated very clearly by the experiments, using very young children, performed by Stumpf. The results can be described in the following manner: the octave is most often heard as a single sound; the fifth as being composed of two notes; and as for major seconds and thirds, children usually perceive them as comprising three sounds, and quite often the major second as comprising four sounds. This is observed as much in younger children who are "musical" as in older children who are not so "musical". These experiments show clearly that children respond to the question, "How many notes do you hear?" by using timbre as their criterion, and not a true auditory analysis. Auditory analysis could not find more notes in a chord than were really there. Now we have seen that with major seconds and thirds children usually hear more than two sounds. Clearly, their judgements are not based on the number of sounds distinguishable by the ear, but on the relative 'fullness' of the resonance.

This fits in very well with the findings described on page 70 suggesting that judgements of this type are influenced to great degree by the position and nature of the chordal items used. The main conclusion from the evidence available is that whilst existing tests of sensitivity to certain aspects of chords and harmony concentrate on the 'How many notes in this chord' aspect, demanding what Teplov calls "l'analyse auditive", most children do not tackle the problem from this point of view, or even hear chords in this way. Rather, what they are conscious of in the transition from single sounds to diads and triads and so on is an increasing sense of richness or fullness in a single sound; not a group of separate sounds. It is also clear that this sense of richness or fullness can be influenced by the choice of material within the chord. (Note that Bentley writes 'So far then, we have found three basic musical abilities that we can attempt to measure in young children - tonal memory, rhythmic memory, and pitch discrimination - and it would appear that these three are indispensable in any musical operation'. (Musical Ability in Children pp 36) and later (39)) 'we have now proposed three abilities as basic, elemental, and essential for music making: tonal memory, rhythmic memory, and pitch discrimination. We have added, as highly desirable, the ability to analyse chords'. So, Bentley sees chord analysis as highly desirable, but not as essential. In view of the fact that much music making places no emphasis at all on this aspect, one is tempted to ask what his criteria for desirability are.)

Let us reconsider the situation in a different light; instead of trying to find out whether subjects can correctly say how many notes there are in a chord, and marking answers right or wrong, try instead merely to discover what normal untrained people, hear. If we derive our standard or norm from experiments oriented in this way there can be no doubt that, for example, a major third seems to have more notes in it than an actave. Such an approach demands that we reconsider our notion of right and wrong answers. Otherwise, if we persist in the old orientation we are merely, to some extent, picking out those children who use "l'analyse auditive" from those who do not. Needless to say, there are a variety of reasons why a child might use the analytic rather than the syncretic method, and these are not necessarily a reliable guide to musical aptitude. Certainly, music teachers when they use their ear tests (some of which involve chord analysis) concentrate on the analytic method; the majority of children without training of this type do not. Thus in respect of this task, trained and untrained children do not start on an equal footing.

Briefly, the arguments against chord analysis can be summarised in the following manner.

1. Children below the age of approximately ten years have no harmonic 'sense', and do not perceive the difference between consonance dissonance. A test using harmonic material is in danger of selecting in favour of earlier developers, rather than in favour of a pure musical criterion.

2. The music of children, and of 'primitive' peoples places emphasis on rhythm and melody. Harmony is either incidental or absent. It is possible therefore that harmony, and particularly the Western conception of it, is very much under the influence of environmental and experiential factors. The wider distribution throughout different cultures, and in children's music, of rhythmic and melodic elements suggests that these are less influenced by specific musical culture patterns. If the above argument is true, then any test of 'western' harmony demands a degree of past experience of this type of sound, if it is to be performed successfully.

3. The normal untrained ear perceives complex sounds differently from the trained ear. The perception of complex sounds is also affected by the nature of the sound source, and by the parts which constitute the complex sounds. The untrained ear, however, is not necessarily an unmusical ear.

On the basis of the above it was therefore decided not to include any material of a harmonic nature, or demanding any type of chord analysis, or any item using consonant (in the widest sense) material. From our original list of twenty-one types of test item, we are now reduced to eight. These are (1) intervals (2) memory (3) rhythm (4) melodic shape (5) creative ability (6) intensity (7) pace (8) pitch.

Further Reduction of Test Material. A Questionnaire.

Two points have already been made in the introduction; that an approach demanding the use of material having musical value would not be adopted here; and that the reactions of musicians to the type of test material would not be an important criterion for test selection. However there is no doubt that in certain circumstances the informed opinions of musicians can be invaluable to an endeavour of this nature.

Before the final selection of test material was made a short questionnaire was devised to try to obtain certain information about musicians' assessments of their own particular skills. Answers were required about certain specific topics central to the construction of a test battery, certain less central topics, and some general introspections. A questionnaire consisting of five short sections was compiled, each part becoming progressively more open ended. A facsimile of the questionnaire is to be found overleaf from page 78.

The group of musicians chosen to fill in the questionnaire was selected in accordance with the following needs:-

1. If the findings were to be most valuable, the sample should comprise only musicians with the very highest degree of ability.
2. A sample of skilled professional performers was desirable, as qualifications in music which did not demand a high level of executive ability tended to produce a sample with less uniformly high abilities.
3. It was desirable that the sample be drawn from a population with teaching experience, as members of such a sample would not be entirely unused to the analytic examination of certain musical tasks. Some degree of breaking down a task into certain simpler elements is helpful in teaching complex musical tasks.

The three criteria listed above were met by drawing a sample from the teaching staff of several well-known music colleges. Many of the members of the teaching staff of such colleges are full time orchestral players, particularly in Manchester and London, where most of the colleges are situated.

Method.

Letters were sent to the principals of all the major music colleges in England and Scotland, briefly describing the area of the projected work, and the nature of the questionnaire. Strict anonymity when refering to contents of individual questionnaires was guaranteed.

Please complete the following questionnaire in accordance with the direction printed below. Where an opinion is called for rather than a factual answer, try to make your answer as objective as possible, and attempt to avoid any purely personal factors which might otherwise influence your judgement.

PART I.

Please supply the following information:-

Name:	
College at which you give instruction:-	
What is your present status as a teacher at the college?	
Instrument on which you give instruction:-	
Are you primarily	(a) a full-time teacher? (b) an orchestral or 'performing' musician, also giving instruction? (tick whichever appropriate)

PART II. Below are given a series of paired alternatives. Each pair of alternatives is concerned with some specific musical ability. You are asked to place a tick opposite one of the alternatives in each pair in accordance with the following instructions:- If you feel that the degree of ability required for the performance of the musical task in which you specialise is on the whole no greater than that required for most other musical tasks, tick the first alternative. If you think that the performance of your musical task requires a greater degree of the ability in question than do most other musical tasks, tick the second alternative. If you feel that the ability in question is not necessary, or is irrelevant to the performance of your particular musical task, put a tick in the 'Rejection Box' which you will find after each pair.

Pair 1.	TICK HERE
Good sense of pitch	
Exceptionally good sense of pitch	
Rejection Box	

Pair 2.	
Good sensitivity to change of intensity	
Exceptionally good sensitivity to change of intensity	
Rejection Box	

Pair 3.	
Good musical memory	
Exceptionally good musical memory	
Rejection Box	

Pair 4.	
Good muscular co-ordination	
Exceptionally good muscular co-ordination	
Rejection Box	

Continued.

Pair 5.	TICK HERE
Good sense of time and rhythm	
Exceptionally good sense of time and rhythm	
Rejection Box	

PART III. Below is given a list of different specific musical capacities. Place a tick opposite those which you feel are necessary for the adequate performance of the musical task in which you specialise.

1. Creative imagination ______ [ ]

2. Above average intellectual level ______ [ ]

3. Musical taste ______ [ ]

4. Interest in listening to music, live or on record ______ [ ]

5. Emotional response to music ______ [ ]

PART IV. Are there any particular physical or bodily characteristics which help in the performance of your particular musical pursuit? (For instance, these might be thin lips, strong hands, long fingers or similar). If you do not think there are, write 'NO', for your answer. If you think there are, please state briefly what you think these are.

PART V. This section is provided for you to fill in any abilities which you think help in the performance of your musical task, and which may not have been included in the questionnaire so far.

Sincere thanks for your co-operation.

The following colleges were approached:-

1. The Trinity College of Music, Mandeville Place, London, W.1.
2. The London College of Music, Gt. Marlborough Street, London, W.1.
3. Royal Manchester College of Music, Oxford Road, M/cr 15.
4. Royal Academy of Music, Marylebone Road, London, N.W.1.
5. Northern School of Music, Oxford Road, M/cr 1.
6. Royal College of Music, South Kensington, London, S.W.7.
7. Guildhall School of Music and Drama, Victoria Embankment, London, E.C.4.
8. Birmingham, College of Music.) no reply.
9. Edinburgh, College of Music.)

From these nine, the following were willing to help in the project:-* The Northern School of Music, the Royal Manchester College of Music, the Trinity College of Music, and the Guildhall School of Music and Drama. Thirty copies of the questionnaire were sent to each of these colleges, except the Guildhall who requested forty copies. 115 replies (completed questionnaires) were received.

Form of the Questionnaire:- Examination of the two previous pages will give the general lay out of the questionnaire. (N.B. Actual questionnaires were printed on two sides of a single sheet.) Briefly, the main points about the form of the questionnaire are as follows:- Part one is purely factual, seeking certain information about the subject. Part Two calls for subjects opinion

* The reactions of some of the colleges who were unable to give their assistance are of interest; however these are outside the scope of this study.

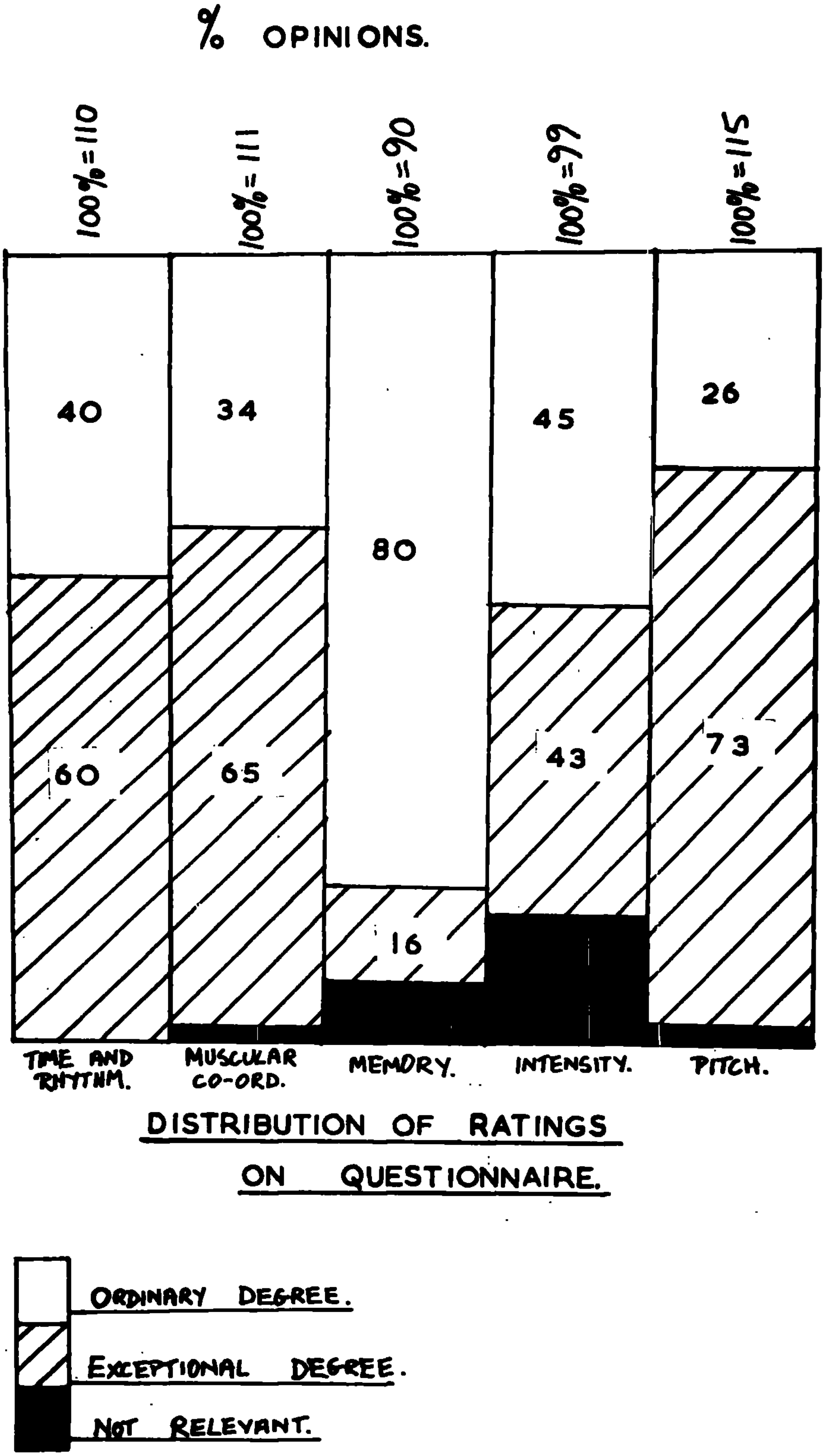
on five specific aspects of musical performance. Part Three calls for the subject to select any number of attributes from a total of five. Part Four asks about physical attributes, but there are no specific questions. The final part is entirely open-ended and allows the subject to raise any matter not contained within the first four sections of the questionnaire. This form was chosen because the answers to certain specific questions were required; and it was felt that a series of more open-ended questions would allow any other significant trends of opinion to emerge. The reason why the instructions to Part Two of the test take the present form are two-fold. Firstly, it was felt that subjects might be encouraged to theorise at length if they were asked questions about general musical performance. It was hoped to reduce the degree of subjectivity in which subjects might be inclined to indulge by keeping questions in this section specific to the musical task in which they specialised, rather than general. Secondly, it was hoped to find which musical attributes were general to all musical tasks, and which specific. By asking for data in this particular form, any general trends can be found even though for individuals the questions are task specific. On the other hand, if questions had been about some more general musical aspect, no specific data would have been obtained.

130 copies of the questionnaire were sent to high grade music teachers (that is, teachers of executive ability rather than teachers of history of music or similar) currently on the staff lists of four high grade music colleges. Numbers sent and institutions concerned have already been described. 115 completed papers were received back; a few individual items on the papers were spoiled, so N = 115 does not hold true for all items on the questionnaire. Totals for items are given in the results.

Results.

The results obtained for Part Two will be given first. These are described in Fig. 1 overleaf. Columns show ratings for pitch, intensity, memory, muscular co-ordination, and time/rhythm. Ratings are expressed as fractions of 100%. The shaded portions show percentage of 'exceptional degree' ratings, unshaded portions show percentage of 'good' (ordinary degree) ratings, and solid block portions show percentage of 'not relevant' ratings. Number of ratings in each column is given at each column head.

Bear in mind that subjects were asked to indicate whether an 'ordinary' degree or an 'exceptional' degree of each attribute



% OPINIONS.
100% = 110
100% = 111
100% = 90
100% = 99
100% = 115
40
34
80
45
26
60
65
16
43
73
TIME AND RHYTHM.
MUSCULAR CO-ORD.
MEMORY.
INTENSITY.
PITCH.
DISTRIBUTION OF RATINGS ON QUESTIONNAIRE.
ORDINARY DEGREE.
EXCEPTIONAL DEGREE.
NOT RELEVANT.

fig 1

was required for the performance of the task in which they specialised. Two things are apparent from the data received on this part of the questionnaire. A significant majority of subjects rated pitch as being required to a greater degree for the performance of their specialist task than for the performance of other specialist tasks. On the other hand, memory seems to be required to a good degree for all the tasks examined, but few subjects claimed their task required it to a greater degree than other tasks. Very few 'not relevant' ratings were obtained.

The table below gives the distribution of results shown in Fig. 1 in terms of z - scores, discounting 'not relevant' responses.

	Z - Score.	
MEMORY	6.4	p = .000000002
PITCH	4.7	p = .0000068
MUSCULAR CO-ORDINATION	3.1	p = .00194
TIME AND RHYTHM	2	p = .0456
INTENSITY	.2	p = .8414

TABLE I.

Distribution of 'Good' and 'Exceptional' Responses.

It can be seen from Table I that the distribution of ratings between 'good degree' and 'exceptional degree' for memory and pitch is extremely significant. The distribution for muscular co-ordination is also highly significant. Taken to the nearest % the distribution of 'not relevant' responses was as follows:- Pitch 1%, Muscular Co-ordination 1%, Time and Rhythm 0%, Memory 4%, and Intensity 12%. By far the greatest percentage of 'not relevant' responses thus fall in the Intensity section.

Part Three: In this section subjects could select, as being necessary to the performance of their specialist task, any of five given alternatives. The percentage of the total subject pool voting for each alternative is given in the table below.

CREATIVE IMAGINATION	52%
ABOVE AVERAGE INTELLECTUAL LEVEL	31%
MUSICAL TASTE	91%
INTEREST IN LISTENING TO MUSIC, LIVE OR ON RECORD	91%
EMOTIONAL RESPONSE TO MUSIC	79%

Questionnaire, Part 3. Distribution of votes (%)

The relationship of musical aptitude to intelligence and creativity will be discussed later, but the results above show that the sample of teachers tested had a certain insight into these relationships.

Part Four: This section was open-ended, and results are shown below. The data can be viewed most usefully in relation to the tasks in which the different subjects specialised. Basically there are three ways of viewing the division between tasks. From the 115 replies, individual job specialisation was as follows:-

Piano'	12
Timpani	8
Violin	11
Viola	4
'Cello	10
Double Bass	10
Clarinet	10
Oboe	6
Bassoon	6
Flute	8
Trumpet	5
Trombone	4
Horn	7
Voice	14

These can be grouped as follows into instrumental 'families'.

STRINGS	Violin	
	Viola	
	'Cello	
	Double Bass	Total 35.
WOODWING	Clarinet (single reeds)	
	Oboe (double reeds)	
	Bassoon (double reeds)	
	Flute	Total 30.
BRASS	Trumpet	
	Trombone	
	Horn	Total 16.
OTHER	Timpani	
	Piano'	
	Voice	Total 34.

On the basis of the type of overall operation called for in the performance of the musical tasks listed, the following general categories can be formed:-

BLOWING INSTRUMENTS.........	Clarinet	
	Oboe	
	Bassoon	
	Flute	
	Trumpet	
	Trombone	
	Horn	Total 46
SCRAPING INSTRUMENTS........	Violin	
	Viola	
	'Cello	
	Double Bass	Total 35.
BANGING INSTRUMENTS.........	Piano	
	Timpani	Total 20.
OTHER......................	Voice	Total 14.

The category 'banging instruments' is less clearly defined than the others. A quick examination of the above will help in interpreting the results which will be presented in accordance with one of the above systems. Results here provided no surprises, and were very much what one would expect. Individual subjects occasionally included material of a bizarre or irrelevant nature, but the following broad trends emerged:-

<u>Blowing Instruments</u> (N = 46)

97% of subjects specialising in blowing instruments stressed the importance of sound, even or normal teeth, and normal lip formation. Mention was frequently made of the disadvantages of abnormally formed top lip, or protruding 'horse' teeth.

<u>Scraping Instruments</u> (N = 35)

Over 90% of performers in this category stressed the importance of both strength and flexibility of hand and more especially fingers. Approximately half of the subjects stressed the advantage of large hands, and half the advantage of broad but not necessarily large hands.

<u>Banging and Other Instruments</u> (N = 34)

No major trend was apparent in this group; perhaps because the classification here is less homogeneous than the previous ones. In general pianists suggested the following:- 30% large hand or good span, with strong fingers. 40% no particular size of hand necessary as 'skill develops to suit the individual type'. Timpanists generally gave no special requirements, but 33% stressed strong supple wrists and good co-ordination of arm movements. Voice specialists (80%) stressed, fairly unanimously, the importance of good breathing apparatus, and 'good, well formed larynx'.

Within the category 'blowing instruments' differences were apparent between the requirements for brass and woodwind players. These centred around finger and hand movement. Normally formed hands and fingers, of average strength were listed in some way by all woodwind players. Brass players made no mention of hand or fingers. The fingering on the woodwind instruments is generally much more complicated than on brass instruments; woodwind instruments call for a unique position of keys for each note; brass instruments demand combinations, none of which are unique or specific to any one note, on (usually) three valves only.* Also, brass fingering involves only one hand; woodwind both.

Part Four: This part was completely open-ended. Certain subjects used this section for brief exposition of personal theories or bias, and much material was not usable. 41% of subjects made no answer in this part of the questionnaire. However, of those supplying answers to this section 30% stressed that no account had been taken in the questionnaire of various motivational factors. Words actually used by the subjects often included 'dedication', 'application' or 'devotion'. In these results, all such answers were taken as being motivationally orientated. Other points were raised by small groups of musicians, which were of interest but not sufficiently well-represented to provide any significant trend.

* Certain brass instruments, like the modern horn, may have four or even five valves. However, two of these are primarily for key change operations, and are often used in an open or closed position for long periods.

Conclusions.

Any conclusions drawn from the foregoing results must be of a tentative nature. Although the overall sample size was 115, the numbers specialising in different specific musical tasks is rather small. However, the questionnaire was intended only as a pointer in the selection of types of material to be included in the proposed test battery.

Part Two of the questionnaire yielded the most interesting results. The number of 'exceptional' votes given to the five attributes placed them in the following order; 1. Pitch. 2. Muscular co-ordination. 3. Time and rhythm. 4. Intensity. 5. Memory. The vital role of pitch in all musical tasks is thus confirmed by the results. Since all existing batteries also include some kind of pitch test we would have the greatest confidence in including some test of this nature. The results concerning memory are rather strange. Although 80% of subjects rated it as necessary to a good degree, only 16% thought it necessary in the 'exceptional' degree. This result was not anticipated. The low number of 'irrelevant' responses suggests that all the attributes included were of some importance at least.

In Part Three of the questionnaire, the percentage of the total subject pool voting for each alternative has already been given. The most interesting results here are those for creative imagination and intellectual level. Of the five alternatives, these two received the fewest number of votes, suggesting that subjects had made to some extent a realistic assessment of the relationship between these and musical ability. The relationship between creativity and scores on perceptual tests of musical ability, and I.Q. and musical ability scores will be discussed in more detail later (Sections 4 and 5); studies have shown that the relationship with I.Q. is generally between .3 and .35. A study of musical creativity and scores on a perceptual test suggest that a similar low correlation exists here.

On the open-ended sections (parts four and five) any conclusions drawn must be of a very general nature, and are more appropriate as guides in selecting suitable musical tasks for certain individuals than for selection of test material. These results are discussed later.

Discussion.

The questionnaire was sent out during the early stages of the investigation, before any conclusions about type of material had been reached. It is necessary to point out initially that much of the material obtained from the questionnaire turned out to be redundant at a later stage. Part Two was the only section which had any direct influence on choice of items for a perceptual test; the remaining parts provided information which has value primarily from the point of view of guidance and selection of instrument. (Noble, 1964, Lamp, C. J. and Keys, N., 1935)

Part Two of the questionnaire contained four items which could be used in a battery of perceptual tests (viz. time and rhythm, memory, intensity, and pitch). Muscular co-ordination was also included as at the time it was hoped to devise a simple test of motor skills to supplement the perceptual material. This attempt was later abandoned, though the high rating given to this aspect by the musicians suggests that such a test would be most useful. The central position taken by pitch in all musical activities has been stressed by many workers. By far the greater proportion of the extant test material contains some measure of pitch. There are a few exceptions (Madison, Gordon and Drake). This is hardly surprising as music consists, apart from rhythmic elements, of structures compounded from noises which differ from one another in frequency, i.e. pitch. Existing tests also suggest the importance of rhythm and some form of musical memory. On the other hand, the importance of intensity has, with the main exception of Seashore, not been stressed. In view of the above we might anticipate that in any questionnaire type survey these things would be reflected. In fact, pitch does receive the highest rating, followed by (of the perceptual items) rhythm, and then intensity. Memory scores very low, which was not anticipated. We would have expected that the relative positions of memory and intensity would be reversed. There are reasons why memory scored so low, based on certain inadequacies in the questionnaire.

The concept of memory, when applied to music, is very general; and insufficient information was given on the questionnaire as to the precise meaning of 'musical memory' in the present context. The sample chosen to fill in the questionnaire consisted entirely of 'performing' musicians; of these, 68% were primarily orchestral musicians also involved in giving instruction. It seems probable that for many of these people, musical memory meant the capacity to 'memorise' and play 'by heart' certain passages of classical music. In general, the only musicians required to do this are solo performers. The ordinary orchestral player always has his music to guide him. It follows therefore that 'memory' of this type is not required in exceptional degree by the majority of players. This situation was reflected in the answers of a small number of subjects who wrote footnotes to the memory question stressing that memory was more important for solo performance than for general orchestral work. One violinist in particular wrote, at the very end of the questionnaire, "About playing from memory, memorizing. Shouldn't memory be listed separately from memorizing?". Clearly, in this instance the musicians' concept of 'musical memory' was not the one intended in the question, and the dilemma is crystallised in the above quotation. Generally, the two types may be distinguished as follows. The musician is apt to interpret 'musical memory' as the capacity to reproduce with complete accuracy lengthy passages of classical music upon a certain instrument, in the absence of any written parts. 'Musical memory' as used in the questionnaire referred to a far less complex task; namely the ability to hold in short term memory the general, overall characteristics of short musical extracts, such as would enable a person to make comparative judgements about two similar, but not identical, extracts, played one after the other. No type of physical reproduction is involved or implied here, least of all any reproduction demanding memory to perfection of long pieces of written music or intricate series of complex executive movements. The type of musical memory intended in the questionnaire is described in the following manner by Bentley; "In order to make accurate response to melody, a child must be able to perceive, and then retain in memory for at least a short period of time, a given order of pitch intervals and note lengths....... When he can remember these in sufficient detail to identify a change in the melody, he has reached a stage of analysis". It is suggested therefore that the unexpectedly low rating given to 'memory' is due in considerable part to the fact that the above definitions were not made apparent in the questionnaire.

Results for the remaining parts of the questionnaire are self explanatory and no further discussion is included here.

Final Selection of Test Items.

We have a short list of eight types of item which merit further consideration for inclusion in the test battery. These are: 1. Intervals; 2. Memory; 3. Rhythm; 4. Melodic Shape; 5. Creative Ability; 6. Intensity; 7. Pace; 8. Pitch.

It has been stated (page 73) that Bentley saw as indispensable tonal memory, rhythmic memory, and pitch discrimination. The importance of some pitch measure has been stressed by other writers also, and most tests include a measure of this type. The questionnaire confirms that musicians also regard this type of ability as being very important. Without the concept of pitch, any kind of melodic development is impossible in music, and we are left with only a bare rhythmic skeleton. In short, pitch variation is what turns rhythms or time patterns into tunes. Some type of pitch measure is therefore selected with the utmost confidence as an essential part of any proposed test battery.

According to Bentley, some type of rhythmic measure and some sort of tonal measure are also necessary. However, in Bentley's system these two are inextricably tied up with memory. All three aspects remain in the short list, but memory is postulated here as a separate entity. More needs to be said therefore about the relationship between certain specific musical capacities and 'memory', either general or specific.

On the basis of observation of children's musical behaviour, and a certain amount of experimental data, Bentley makes a distinction between two types of musical memory, one of which is specific to tonal material and one to rhythmic material. This distinction is reflected in his test battery: and no other battery makes the same distinction. Most indeed have a separate 'memory' section. Bentley's distinction is a valid one, but does not unfortunately suggest any basically new approach to problems of measurement for the following reasons. Consider some of the existing test batteries; Seashore used tests of memory in which a subject had to detect a change of one note in two consecutive playings of a tonal sequence, and rhythm tests in which subjects had to make comparisons of two tapped time patterns on a 'same' or 'different' basis. Wing included a memory test demanding detection of change of a single tone in two consecutive playings. The Kwalwasser-Dykema uses rhythm tests of, again, two rhythmic patterns demanding 'same' or 'different' response. Whistler and Thorpe devised tests called Melody Recognition and Rhythm Recognition. What form do they take? Once again the comparison of pairs of identical or almost identical pairs of stimuli calling for the response 'same' or 'different'. This 'comparison of pairs' method is very common in the test material available. Now consider the Bentley Tests. His 'Tonal Memory' items use the 'comparison of pairs' method, and so do his 'Rhythmic Memory' tests, though these include a counting element as well. The distinction between two types of musical memory is merely reflected in Bentley's system as a change of name. Examination in detail of the material contained in the systems mentioned above will show that writers have almost always tended to refer to what Bentley calls 'tonal memory' as simply memory; and 'rhythmic memory' as simply rhythm. Thus Bentley's attempt to break down the memory aspect into two parts unfortunately does not lead to any major rethinking with regard to test construction.

On the distinction between tonal memory and rhythmic memory, Bentley writes the following: "Memory for melody, then, is an ability that develops at an early age; and this ability can be measured. When we come to devise means of measuring this ability we recall that the two fundamental characteristics of melody that make one tune distonguishable from any other tune are the tonal and rhythmic aspects. We also note that the errors of detail that we find children spontaneously correcting in their performances are errors in both tonal configuration and in rhythm. The rhythmic aspect frequently seems to cause less trouble than the tonal; but when the rhythm is right children concentrate entirely for the moment on the tonal aspect. Similarly, when the tonal aspect is right they devote their whole attention to the rhythmic correction. They distinguish these two aspects, and deal with them separately."

"This would seem to indicate that separate measurements should be made for these two essential and distinct aspects of memory for melody". Although the distinction has perhaps never been made in this form before, a great many existing tests do in fact measure these two 'distinct aspects' though under a different name.

The above discussion provides an insight into certain problems associated with musical memory in the present context, as follows: Bentley showed that in solving problems of melody and rhythm two types of memory were distinguishable. No mention is made of memory in the chord analysis or pitch tests, and yet these also involve a degree of memory. The problem centres around types of memory which are postulated as being either

(a) specific in some degree to certain types of musical task or (b) more general. The very nature of perceptual group tests ensures that these two are often inextricably bound together. Bentley has produced evidence showing differences in memory for melody and for rhythm. Evidence is also given in a later section (section 5), showing that memory for rhythmic patterns can be differentiated from memory for other types of material. In the testing situation however, the very nature of the test means that all its parts demand short term memory; the subject hears something and has to store it in order to make a comparison with what he hears next. In fact the memory components stressed by Bentley are not the only ones present in the testing situation. Tests such as Bentley's pitch test demand storage of the first stimulus tone in order to compare the second with it. Similarly for Wing's Pitch Change Test. Even chord analysis as envisaged by Bentley and Wing demands storage of the harmonic material while the analytic process takes place. (The point has already been made that if a subject responds to the overall sound in a non-analytic manner, his answer is likely to be in error.) The point being made here is that all aspects are in fact memory tests of some kind in that storage and then recall of material, of different types, is demanded. And by saying that certain types of memory are specific we mean simply that people store different types of material with different degrees of accuracy.

In the present thesis, the relationship between 'memory factors' and musical ability is envisaged as being rather special, and the following section is devoted to this:-

'Memory' and Musical Ability.

In this section it is proposed only to give the very barest outline of a theory of musical ability based almost entirely on certain memory factors, namely processes of encoding and decoding, and recall, of certain types of material. Considerable space is given to a discussion of this topic in section 3, where it is more appropriately situated. We have stated that all the test material available calls for the storage in short term memory, and manipulation of, certain types of material. The implication here is that what is being measured in these tests is just this ability (i.e. storage and manipulation) and that what is meant by musical ability, as far as tests are concerned, is just these things. It has been hypothesised earlier that this in fact comes close to being a useful and operational definition of musical ability in more general terms, but this discussion need not concern us here. If we accept that all the existing test material is concerned basically with memory processes vis a vis certain types of material, it becomes clear that to give certain tests the title of simply 'Memory' in an all inclusive manner, and to describe the other tests as though they were measures of something else is illogical. Similarly, to devise tests of 'Melodic Memory' or 'Rhythmic Memory', and then to complete the battery with the implicit assumption that memory is somehow not pertinent in the remaining items, is also illogical. It also follows, if we accept the argument, that in devising test material we can manipulate the order of difficulty by altering the ease with which material can be remembered; that is by altering the length, and the codability, of items. (The effects of manipulating these variables is described in Section 3.)

On the implicit understanding, then, that all test items proposed here are regarded as being measures of coding and recall processes, all items proposed will simply be given the name of the specific musical task in which context these processes are being assessed. No separate 'memory' test can be proposed as it permeates all items used; the test relating to coding/decoding and recall for pitch material will simply be called 'Pitch', and so on.

On the basis of the above, item (2) on the short list, memory, can be discarded; not because it is irrelevant, but because it is central to all material used, (e.g. Wing's 'memory' test, for instance, would probably be referred to as 'melody' in the present system, since it measures memory for melodic material).

Many of the existing test batteries use some type of test of melodic material. The Wing 'memory' tests have already been described; Seashore also used the method involving two nearly similar playings, but his material was selected 'to form no melodic line'. (But what differentiates a 'melodic' from an 'unmelodic' line is not clear.) Kwalwasser-Dykema (tonal memory), Gaston (melodic memory), also used a similar method of assessment. When discussing children's music and the variables by which it is characterised, the prominent role of melody was stressed. Also, the strong role of melody extends to many cultures, and is a world-wide feature of music. These points have already been made in detail. On the basis of the above, some test of melody was selected with confidence as being an essential part of the proposed test battery. From the short list of eight types of material, 'pitch' and 'melodic shape' have thus far been selected, and 'memory' has been deleted for the reasons given in the previous paragraph. The remaining items are (1) intervals (3) rhythm (5) creative ability

(6) intensity (7) pace.

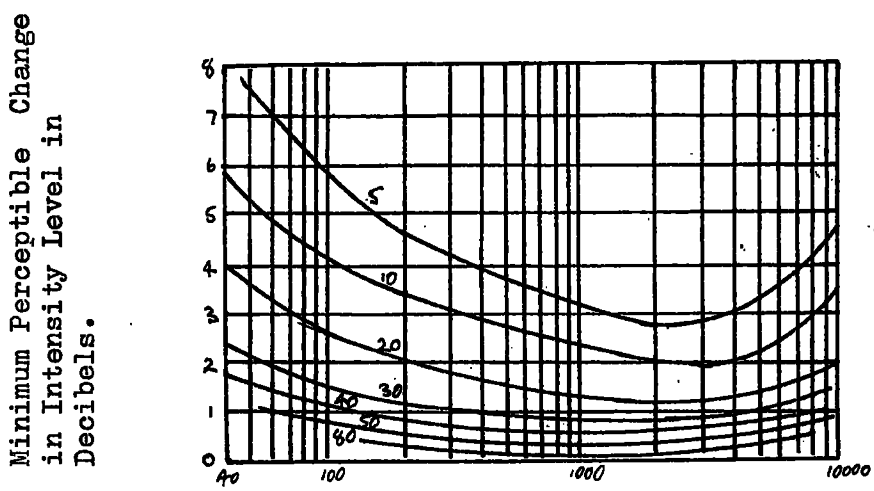
The central role of rhythm from a cross-cultural point of view has also been stressed. Many existing test batteries include some test of certain aspects of rhythmic abilities. Together with melody, rhythm is one of the foundation stones on which all tunes, from the simplest to the most complex, are built. Thackrey's studies have indicated that rhythmic abilities are complex, as reflected in his battery of seven different tests of rhythmic aptitudes. Though the present battery could not absorb such a detailed examination due to pressure on time for administration, some test of basic rhythmic capacities is necessary, and the one finally chosen here is one not included in the Thackray battery.

Creative ability is unsuitable as test material. Though studies of creativity can be scientifically orientated, the major problem of criteria remains unsolved. What in fact is musically creative and what non-creative? Either we have to assume that something else is a reliable indicator of creativity, such as number or uniqueness (Getzels and Jackson, 1962; Wallach and Kegan, 1965) of responses, or take note of informed opinion as to what is creative. The other major problem in the field of musical creativity arises out of the need for subjects to have some familiarity with musical material in order to be able to manipulate it and display their creative talents.

Details are given in a later section of certain experiments in devising tests of musical creativity; one of the findings from these was that scores on the creative tests did not correlate highly with scores using the types of perceptual test material so far described. It would seem therefore that creativity, as measured by certain types of test, is an entirely different kind of ability from the abilities measured by the other test material so far proposed. It is not therefore suitable for inclusion in the present battery.

Of the remaining items, intensity and pace are also rejected as unsatisfactory. Intensity has been used as a measure by some workers, most notably by Seashore. Wing also used a test of intensity, as an appreciation measure, (i.e. "Judging the more appropriate mode of varying loudness..... in two performances of the same melody"). These two approaches illustrate perfectly the dilemma arising from the use of tests of intensity. In the first place, Seashore's test, which takes intensity out of the musical context, requires subjects to state which of two buzzer noises ('first' or 'second') is the louder. This approach has been justly criticised on the grounds that it is a purely psychophysical test and has little to do with musical ability. This is reflected in the low validity of this item. (First Version, validity of intensity test .13, with music grades and teacher's ratings. In the data on the revised version, validy data available only for Pitch, Rhythm and Tonal Memory. Validity for other items described as 'questionable'. Shuter, Appendix 1, p 281). A general conclusion would be that this approach is a test of sensory capacity rather than musical ability. On the other hand, Wing's approach is equally unsuitable for inclusion in the proposed battery. In the musical situation, variation of intensity takes place in a purely aesthetic manner. Sensitivity to changes of intensity in a psychophysical sense is not nearly so important as when and how the changes are made; but when and how changes are made is very much a stylistic consideration, influenced by certain conventions. Any test of dynamics (intensity) in the musical situation is based on these considerations and can only be a test of appreciation, depending on the familiarity of the subject with the conventions.

A closer examination of the role of intensity in music will help to clarify the situation. Variations of intensity are used primarily to convey emotional states in music. Typically, loud passages tend perhaps to convey anger, or ebullience or more overt emotional states; quiet passages convey sadness, wistfulness and so on. The actual changes in intensity, in terms of sound pressure, are however readily discernable. If any part of this eludes the listener, it is the subtlety in the use of intensity change and not the actual change itself which is almost invariably well within his physiological capacity. Below, is reproduced a table showing the minimum perceptible change in intensity level of pure tones as a function of frequency. (Fletcher, 1953).



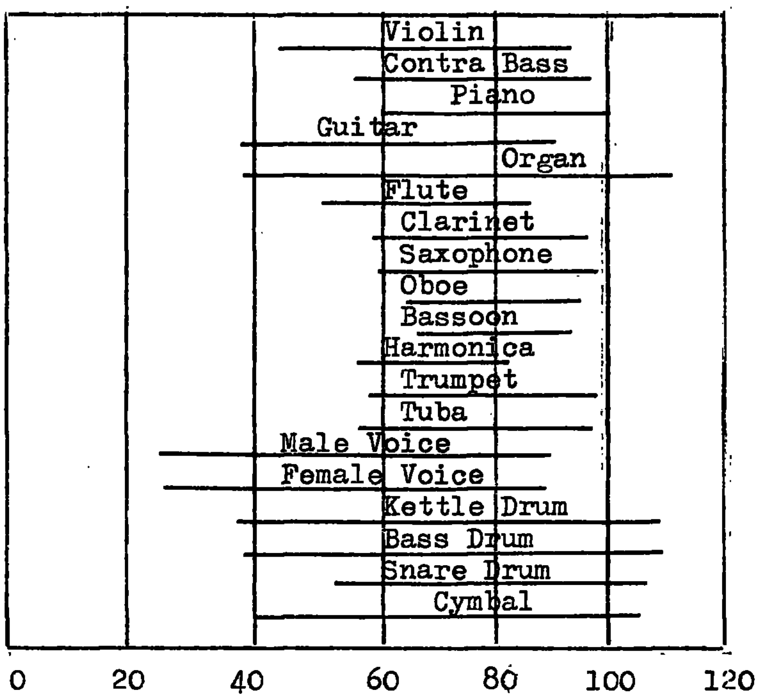
Frequency in Cycles per Second.

Numbers on curves indicate level above threshold.

It will be noted that sensitivity depends on the frequency of the tones, and that sensitivity is greatest around 3,000 c.p.s.; also, there is a differential effect of level above absolute threshold (usually taken as .0002 dynes per sq. cm.) Note that under the most unfavourable condition (i.e. 40 c.p.s. 5d. b.

above threshold) the minimum perceptible change is 8d. b.

A table giving the intensity ranges of various musical instruments under controlled conditions is given below, and serves for interesting comparison (Olson, H. F., 1967), with the previous figure (page 104).



The intensity ranges for various musical instruments at a distance of 10 feet. 0 decibels = 0.000204 dyne per square centimeter, or 10 - 16 watt per square centimeter.

Unfortunately no really reliable data is available on the range of intensities, measured at the ear-drum, produced in various musical performances. There can be no doubt, however, that the intensity differences which are used in music to give certain artistic effects are greatly in excess of the figures given for minimum perceptible change. For example, in a 'normal room' with no individual sound sources discernable (such that the room would be regarded as 'silent' in a non-scientific sense) a sound level meter would probably register between 30 and 40 decibels. The sound of a piece of paper dropped in such a room would register an increase of something in the order of 12 d.b. /ft. from the meter. From the graphs given on page 105 we can observe that for sound level of 40 d.b., the minimum perceptible change in decibels is between 2 and 3. Measurement of threshold sensitivity is likely to be unreliable as a guide to musical ability therefore. The low validity of Seashore's intensity test confirms this.

Finally on this point we may note the following, from Seashore: "There are two measures which are basic to all dynamic* aspects of tone, sensitivity and discrimination. The first is the measure of the natural capacity of the ear for becoming aware of sounds; the second is a measure of the capacity of the ear for hearing differences and, therefore, the power to use the ear in a musically significant way dynamically, that is, to assign musical meaning to loudness characteristics". Seashore's test is based on this premise, with the resulting difficulties described above. The

* In the musical context the meaning of 'dynamic' is quite specific, and refers to any variation in intensity. In such written music, there are indications about the loudness or softness of particular phrases or passages. These indications are known as the 'dynamics'.

fact is that we can now see that measuring "the capacity of the ear for hearing differences" does not imply, as Seashore suggests, that we can say anything about the ability "to assign musical meaning to loudness characteristics". On the other hand the approach of Wing produces tests of appreciation rather than aptitude, with difficulties described earlier. (pages 13, 56 - 57).

Pace was rejected on the grounds that as an ability it was insufficiently central as a component of musical ability to merit its inclusion in a test battery where time and space were at a premium. Also, pace is essentially a temporal element and as such can be regarded as an aspect of rhythm. It has already been stated that the rhythm test finally derived in this work is not intended to be an exhaustive coverage of rhythmic abilities. A comprehensive battery of rhythm tests has been devised by Thackray (1969). In this battery the notion of pace is referred to as 'tempo'. Pairs of stimuli are presented and subjects are asked to indicate whether the 'tempo' of the paired items is the same, or if not, which one was quicker. In a factor analysis of the tests he devised, Thackray found that the most satisfactory test to use, where maximum predictability was required from the use of only one type of test, was one in which the subject was asked to reproduce the rhythmic pattern from a piece of heard music. This is hardly surprising since a task of this nature is complex and to some extent involves all the separate tasks in his battery. However, the use of melodic material in tests of rhythmic ability has certain drawbacks; these are discussed in Section 3, when the proposed rhythmic test is discussed in detail. It is interesting to note that in a paper given by Thackray to the Annual B.P.S. meeting, Education Section (Music), September, 1968, he described 'Tests of Rhythmic Ability'. This differed in some respects from the battery described in Music Education Research Papers Number 4 (Novello, 1969); no test of 'tempo' was described, and a test of 'steadiness' was included, which does not appear in the latter.

Conclusions from the above, as far as the production of rhythmic test material is concerned, are that where it is possible to use only a single test then best results will be obtained from items of a complex rather than a simple nature. The exact nature of various rhythmic tasks will be discussed later, but in conclusion we can say that it is desirable to devise a test demanding the apperception of several variables including pace or 'tempo'. In the present battery, this is achieved by producing tasks demanding a type of pattern recognition in which the subject has to extract and compare information from two different types of source. No test of this type is contained in the Thackray Tests of Rhythmic Aptitude.

In summary, examination of certain existing data leads to the conclusion that in the present battery no separate test of pace is warranted. It is possible nevertheless to devise rhythm tests of such a nature that good predictive value can be obtained from test material of a complex nature. Unfortunately no validity data is yet available for the Thackray Tests, but the nature of the items suggests that complex tasks will have higher validity than the simpler ones included in his battery.

The single remaining item from the original short list is interval. In the development of his test battery, Herbert Wing tried two interval tests, one in which subjects had to name the intervals, and one in which two intervals were played and a 'same/different' type of answer was required. Of these two tests he writes (selection of the most suitable items, pp 44 - 45), "The test on stating the names of intervals,......., gave a high correlation with the total scores, but that on comparing intervals,.... which was designed as an easier variant of the same aspect, showed very little agreement". The first of the two forms, clearly is unsatisfactory since formal musical knowledge of interval names is demanded. The second form Wing rejected because average scores were so low as to exceed guessing level on only four items. It was also shown that in the performance of these tests musical subjects often gave the wrong answer due to an inability to consider the two stimulus intervals independently, and instead they often compared one of the intervals with the interval between the two paired stimuli*. Similar difficulties are reported by Heinlein, H. (1925) in a study of the Seashore Consonance Test. Tests of interval discrimination were also used in a test battery by Madison (1942). Separate studies showed reliability of .74; .76, and .84.

Notwithstanding the difficulties that musical subjects found in understanding the nature of the task in Wing's interval tests, Wing also wrote, "When the subject was told to analyse the first interval and then the second, and finally to write down whether the two were the same or different, the musician was undoubtedly superior to the unmusical person in this task". In other words, the tests discriminated between 'musicians' and 'unmusical persons' but the discrimination was based, at least to some extent, on formal musical knowledge needed for the 'analysis'

* It has been pointed out that much test material has been devised from the point of view of the musician. Of the problem described here, Wing writes, "When the subject was told to analyse the first interval, and then the second, and finally write down whether the two were the same or different, the musician was undoubtedly superior to the unmusical person in this task. However, this is a departure from the normal habits of the musician in listening to music. Obviously, any test which penalizes the subject who follows good habits of listening to music must be discarded". This passage is a logical outcome of "..... the standpoint of this investigation is nearer that of the musician than that of the physicist.....", but illustrates the dangers of such an approach. If we select all our items so that they accord nicely with 'the subject

of the items. It will be argued here that it is possible to arrange an interval measure in such a way that no formal analysis is necessary, so that untutored but not necessarily unmusical persons are enabled to score highly on the tests; and to arrange instructions so that musicians are quite clear as to what is required.

The evidence from previous studies is not sufficient therefore to rule out the possibility of devising workable tests of interval. There are reasons to suggest that an interval measure is an essential component of any test of musical aptitude. The manner in which interval material has been used in previous test batteries can be viewed as an extension of harmonic test material. In the present study, it is an extension of melody. This difference arises out of the fact that in previous batteries the sounds comprising the intervals have been played simultaneously, in the form of a chord. In the proposed battery the sounds comprising the interval are played separately, one after the other. Whereas the first approach presents interval as an element of

* Continued from overleaf.

who follows good habits of listening' we penalise the person who does not listen. Clearly, we do not want to select against the person who has an experiential advantage; but we must not select for him either. The fairest way of doing this is to select material with which he is as unfamiliar as the person without any experiential advantage.

harmony, the second presents it as an element of melody. We have already examined the reasons for not using any harmonic material; and described how, in Stumpf's experiments, the number of notes heard in a diad depends on the interval between the constituents. It is clearly meaningless to ask persons to compare the differences between intervals when the normal ear hears different numbers of tones in the different intervals. (The experimenter is in effect asking 'how far apart are these two notes' when the subject hears perhaps three or four tones.) Also, 'l'analyse auditive' is called for whenever subjects are asked to attend to individual tones in complex harmonic structures. We have also noted how there is no cross-cultural basis for harmonic material, and how it is a feature absent from the music of children in their spontaneous music making. (These arguments are expounded in detail on pages 63 - 76) The above arguments do not apply when interval is taken as an element of melody. To distinguish it from previous treatments of interval, the present interval measure will be referred to as 'auditory transposition'. The importance of

a test of 'melodic shape' has already been stressed. Any 'melodic shape' is characterised by tones of certain frequency (pitch) situated at different points of the frequency spectrum, that is notes separated by intervals. There may or may not be a rhythmic element, but a test of this has already been included. It is important to show that interval is something different from melodic shape and pitch, however, if redundancy is to be avoided. This difference can most easily be grasped if we look at these three 'tonal' abilities (as distinct from rhythmic) in the following manner. To perform pitch tests we have to listen to a standard frequency, store this at least for a short time, and be able to respond by recognising tones of the same frequency when they next occur. This also implies that we can tell when subsequent tones are of a different frequency. Melody or melodic shape calls for a response to a series of different pitches, and involves the storage, at least for a short time, of the relative frequencies of the constituent tones. Note, however, that absolute accuracy with individual tones is not essential, nor even aesthetically desirable. Provided the main relationships are grasped, the melodic shape can be said to have been retained. Interval, or 'Auditory Transposition', is not a measure of retention of individual tones, nor a measure of retention of longer series of overall relationships. It involves storage and recognition on the basis of the <u>frequency ratio</u> of two tones.

It is true that we can describe any fragment of melody in terms of the intervals between the constituent tones; a melody (apart from any rhythmic aspects) can be viewed as a succession of intervals. It might appear therefore that by proposing a test of interval we are merely duplicating the 'melodic shape' test on a more elemental level. There is evidence to show that perception of interval is quite different from perception of melody, however. Firstly, memory for melodic shape seems to be quite different from memory for intervals. Experiments by the writer have shown that in testing memory for melodic material, provided that the degree of randomness or organisation of the melodic material is kept constant, then longer passages are more difficult to remember than shorter passages. With progressive shortening of items, more accurate responses are obtained, until the material has been shortened down to two notes only. An increase in difficulty is observed when this stage is reached. This is not due simply to key change factors either. Subjects can recognise tunes they know regardless of what key they hear it in. The memory is for the pattern of relationships and not specific to the key in which they first heard the material. Thus most people can recognise 'God Save the Queen' regardless of starting note, even though they will almost always have heard it in the key of G major. Memory for interval then cannot be explained as simply a form of memory for very short bits of melodic material. Teplov distinguishes two aspects in the correct perception of melodic material (as distinct from melodic shape) which are: 1. that in which all that is perceived is "la courbe melodique", that is to say that the only thing perceived and stored correctly is the direction of movement of the constituent sounds, "la succession des montées et des descentes". 2. not only the direction of movement is perceived and retained correctly, but also the actual size of the interval steps. Meissner performed experiments with 700 children aged 8 - 14 years. They were asked to reproduce a melodic extract which they had never heard before, after a single presentation. Of those who failed to reproduce correctly (which was the majority) over 50% correctly reproduced the 'melodic shape', that is what Teplov describes as "la courbe melodique", but failed in the accurate reproduction of the intervals. (Meissner, 1914) The difference between perception of melodic shape and interval has also been shown by Brehmer (1925), and Stern (1927). Gesell and Ilg (1943) also reported how children aged 3 years could reproduce whole songs recognisably, though not in pitch (i.e. the intervals were not rendered correctly). It seems from the available evidence that the correct perception and retention of complete melodic material hinges first of all on a grasp of the general shape of the melody, and a later stage when the exact separations of the various intervals are perceived and retained.

If we accept the importance of melody in music, and wish to devise satisfactory tests then both aspects mentioned above (melodic shape and interval) must be included, since the absence of either will hinder the development of general melodic sense. On the basis of the above evidence it was concluded that a test of interval was an essential part of any proposed test battery.

From the final short list the items selected as most essential are:-

1. Pitch.
2. Melody.
3. Rhythm.
4. Interval.

All other items have been examined and rejected on the grounds that they were unsuitable, or if suitable not sufficiently important to warrant inclusion in a test battery of limited length.

SECTION 3.

Development of Test Material.

Some Additional Considerations in Test Construction.

Regardless of whether one takes the 'atomistic' or the 'unitary' view, one can observe that musical ability manifests itself in a great many ways; and no single act can be said to be the total manifestation of musical ability since all demand some degree of specialisation. Three broad categories of musical behaviour are commonly cited (Wing, Bentley and Seashore all make the same categorisation). These are listening (in a 'musical' fashion), performing, and composing. Bentley says, "The person who composes music may be safely regarded as musical, even if there are differences of opinion about the quality of his compositions. The performer who never composes may also be safely regarded as musical, although in a different way; he recreates in sound the ideas the composer has imaged and recorded in the score by means of visual symbols. The listener who neither composes nor performs may also be a musical person; for the composers ideas, recreated in sound by the performer, still have no meaning until they have been heard and understood by the attentive listener".

"All three, composer, performer and attentive listener, are 'musical'; all three possess characteristics that distinguish them from those who neither compose, nor perform, nor listen to music".

Generally speaking, writers agree that the composer possesses the highest degree of musical ability, followed by the performer, and lastly the attentive listener. The above implies that there are differences of some kind between the three types of musical person. Clearly there is a difference between performing skill and composing skill. The position of the attentive listener is less clear cut, though listening is a basic requirement for all musical development. However, it is not the differences between these three types of musical activity which are important; the most striking things are the similarities. It is suggested here that the 'complexity' of 'musical ability' stems in no way from the variety of gross musical tasks available, but from the relatively few psychological tasks which are common to musical activities of all kinds. 'Musical ability', whatever it may be, is essentially a mental ability; and its complexity stems only from the variety of purely mental processes involved. (For instance, the muscular co-ordination necessary for handling a violin bow is only pertinent to the development of violin playing, and not directly to the development of musical ability, except through facilitation.) In other words, given that a composer has a very high degree of 'musical ability', the best performer will be the one who, in addition to perfect technique, has the same musical abilities as the composer. Similarly the listener who is able to 'appreciate' and 'understand' a piece of music will possess the same mental musical abilities as the performer and the composer, though perhaps to a lesser degree. Note that almost without exception, all performers can and do compose to some extent, most composers perform, and all are attentive listeners. From this we may conclude that the differences of kind between them stem not from differences in the type of musical abilities, but from differences in skills resulting from specialisation.

This view is directly opposed to Seashore's statement, "As we have seen, musical talent is not one but a group of hierarchies of talent. The musical person may be distinguished in voice, in instrumental performance, in musical appreciation, or in composition; each of these is an independent field in which one may gain eminence without giving evidence of marked ability in the others". If such a view were wholly tenable, to use any single battery of tests of musical ability for assessment of general musical potential would be impossible. It is the fact that the mental abilities underlying various musical tasks are broadly similar that makes 'musical ability' measurable.

Musical ability manifests itself in a variety of behavioural and mental responses. Each separate type of behavioural response depends on a combination of acquired skill and natural aptitude, which combine to produce a certain ability level in the performance of that response. These responses can be categorised on the basis of the operations necessary for their adequate performance. However, if a particular <u>mental</u> response is, in a particular case, deficient to a significant degree, we can say with some confidence that 'musical ability' will not be present to a high degree. By this, we mean only that the absence of the particular mental ability may prohibit to some extent the adequate performance of musical tasks which are the truest evidence for musical ability, since a minimum level of proficiency in that task is essential to the performance of the complex musical task as a whole. In measuring these things, however, we need not concern ourselves with the problem of whether 'musical ability' is in fact greater than the sum of the parts, requiring that it be postulated as a separate entity permeating all the individual parts; or whether it is merely a term to describe the sum of the parts. If sampling certain of the parts gives a useful guide in terms of some musical criterion, then the study of the parts is all that is necessary. It is not necessary to put all one's eggs in the same basket, and follow either the atomism of Seashore or the unitarianism of Wing, in order to draw up a test battery.

NATURE OF THE TEST MATERIAL.

Many existing test batteries use material of an exclusively musical nature. The argument against this approach is, briefly, that questions or problems based on formally musical material are influenced by the subject's familiarity with such material. Herbert Wing writes of his own tests, "....the standpoint of this investigation is nearer that of the musician than that of the physicist....". He also stresses that any proposed tests "must be acceptable in their basic principles to musicians", and makes the point that tests which do not have intrinsic musical value are unlikely to be adopted by music teachers. This point of view is not the one taken in the present study. The researcher in this field must have a completely free hand, at least in the initial stages, to produce test material with validity and reliability as the only ends in view. Possible lines of investigation must not be passed over simply because music teachers find the items dull or uninteresting, or because they do not find a certain type of analysis attractive. Note also that, from the point of view of testing, there is no reason for supposing that tests of musical ability or aptitude should necessarily be composed of musical material. Anastasi writes, (1961), "It should be noted in this connection that the test items need not resemble closely the behaviour the test is to predict. It is only necessary that an empirical correspondence be demonstrated between the two". And later, "It is entirely possible, for example, to devise a test for predicting how well an individual can learn French before he has even begun the study of French. Such a test would involve a sample of the types of behaviour required to learn the new language, but would in itself presuppose no knowledge of French. It could then be said that this test measures the individuals "capacity" or "potentiality" for learning French".

Compare the above with Mursell's statement, "Only the observations of the subject in various musical situations are a guide to the degree to which talent is present".

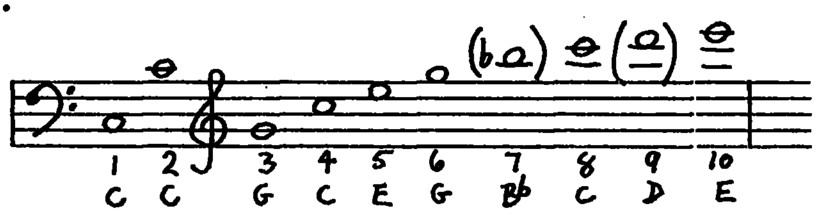
In this context, we might make a comparison between tests of musical "potential" and Anastasi's suggested test of French "potential". The futility of measuring potential for learning French by presenting a subject with a test, all the items of which are in French, is easily seen.

There is no attempt, therefore, in the present study, to produce items with any musical value whatsoever. Since the use of such material carries a heavy penalty in the form of experiential or environmental bias, it was decided that no such material be used here; the material used is best described as quasi-musical.

The problem of medium of presentation also presents itself. Two broad methods are available for the production of musical sounds; one can use orchestral instruments or recordings of them; or electronic means of production can be employed. There are several facts suggesting that the electronic method is likely to prove most satisfactory.

Teplov, reporting on experiments by Meissner (1914) and by Stumpf (1883) shows that familiarity with the 'timbre' or tone quality of the pianoforte is a crucial factor in the performance of certain tasks, particularly chord analysis tests of the type later devised by Wing. All instruments have a characteristic tone quality (timbre), and the person familiar with this sound is at a distinct advantage. (Teplov) In addition to such cultural considerations, there are other aspects of using musical instruments, and particularly instruments like the piano', which make them unsuitable for the present purpose.

The sound response produced by a piano is designed to give the most aesthetically pleasing sound, rather than the purest or most 'simple' one. To this end the hammers are so arranged that they strike the strings, on the majority of instruments, at a point about one seventh of the length of the string from the end. Harmonics are produced in a certain arrangement by this means; in particular the first two inharmonic partials disappear (Helmholtz). These are the seventh and ninth partials. The seventh approximates to a minor seventh, and the ninth to a major second, relative to the prime (fundamental). Clearly, other inharmonic partials do occur above the ninth partial, but these become progressively fainter, so that their overall effect on the nature of the sound may be considered negligible. The following diagram shows the mode of occurrence of the partials, for a tone containing all partials. The first ten partials only are given, as any remaining ones are likely to be very weak. The fundamental is given as a 'C' one octave below middle 'C'.



The above shows all harmonics from a struck string, up to the number ten. In pianos, the string is so sounded that certain partials are weakened so as to be virtually absent, in the following manner. When a string is struck, partial tones are produced by wave forms set up within different sections of the string; however, all those wave forms will be damped which have nodes at the point at which the string is struck. This is because the string is displaced at a point at which it would have to remain stationary if certain wave-forms were to be produced. By striking the string at a point one seventh of its length, the seventh harmonic, formed from vibrating sections one seventh the length of the string, is effectively damped out.

The ninth partial is also considerably weakened. (Seventh and ninth partials are in brackets in the above diagram) The described arrangement of harmonics gives the piano its characteristic sound. In a similar way, all other instruments of the orchestra have their own individual sounds by virtue of a unique arrangement of harmonics; and the final 'tone colour' of the instrument is a product of the arrangement of harmonics and also their relative intensities. In the case of the piano, for example, the upper partials can be rendered more, or less, powerful by using harder or softer hammers. Note also that the relative intensities of the various harmonics are not constant throughout the frequency range of the instrument. For example, a bassoon playing 'C' 523 c.p.s. has 87% of its energy in the first partial. For the note 'E' 163 c.p.s. there is <u>no</u> energy in the first partial (i.e. this note has no produced fundamental) but the third partial has 87% of the energy. In the piano this situation is further complicated by the fact that every note is represented by a different set of strings, and there are three strings per note. There is no guarantee that the response characteristics of all the strings are the same; and taking into account changes of material and of diameter, this is most unlikely.

The main point is that all these instruments produce sounds which are compounded from different tones. The fact that these are phase-linked does not guarantee that a single percept will result in the case of a particular subject. Conversely, where tones are not phase-linked, there is no guarantee that a single percept will not be heard. Helmholtz (1885) showed that with practice it was possible to hear particular partials in phase-linked musical sounds, (in the case of tones of very low frequency, such as those produced by a drum, the pitch of the fundamental can often only be inferred from the harmonics) and Stumpf (1898) showed how chords (sounds composed of several non-phase-linked tones) could be perceived as single percepts. The above observations are clearly relevant to any musical testing procedure, and particularly for tests of harmony or consonance. For example, consider the case where three tones are struck on the piano', and the subject is asked to say how many notes are present. (We have already noted how choice of constituent tones influences a subject's judgements) The test constructor makes the decision that 'three' is the correct answer, because he has played three notes. In fact, if we viewed on an oscillascope the wave form that results from his pressing down of three piano' keys, we could well find that upwards of thirty tones were present. By using pure (sine-wave) tones in this situation the number of tones present can be drastically reduced. In fact, we can never reach the situation where only three tones are present, since difference tones, aural harmonics, summational tones, and sounds due to sympathetic resonance in the testing room, can never be eliminated in a group-testing situation. They can nevertheless be reduced by using sine-wave tones.

Two reasons for not using musical instruments have been put forward. Namely, that subjects familiar with the sound of the instrument used have an advantage over those who are unfamiliar with the sound. By using sine-wave tones we use a sound with which both musical and unmusical groups are unfamiliar. There is no reason for supposing that musical children are, in the main, any more familiar with sine-wave than unmusical children. Secondly, the number of partials present in the complex tones of musical instruments raises problems of measurement in chord-analysis and pitch experiments (where a subject might match a tone with one of the stronger partials). In these situations, it is impossible to be sure whether the response criterion of the subject is the one intended by the experimenter.

A much more important consideration with respect to complex tones is that a subject's ability to perform certain tasks is a function of the complexity of the sounds used (Slawson, A. W., 1968; Schouten, J. F., 1968). Studies have been performed showing that, with complexity held constant, complex tones are more easily matched than sinusoidal tones; and that octave errors are very common when two different types of periodic complex sounds are matched. (Ritsma, 1966) The upper partials of complex tones provide extra cues that are not present in sinusoids, which makes them easier to match provided they are both of the same complexity. Where tones are of different complexity, matching is more difficult. A series of experiments was performed by the present writer in an attempt to clarify this situation in relation to 'musical' and 'non-musical' groups.

Experiments in Pitch Matching.

Introduction.

In view of certain evidence on the effects of 'timbre' on pitch judgements, it was desirable that the exact influence of 'timbre' in certain relevant situations be investigated. Many existing test batteries have 'timbre' as a largely uncontrolled variable. For example, the use of pianos in pitch matching experiments means that the 'timbre' of the different tones presented to the subject is constantly changing.

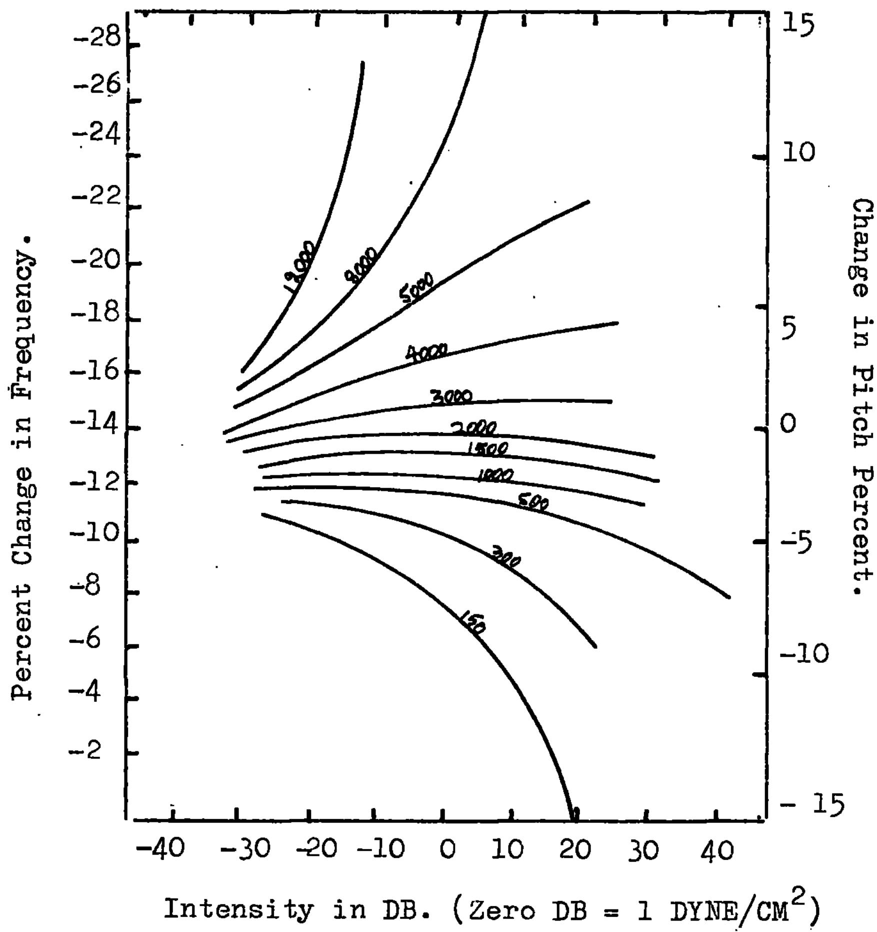
In addition to measuring the effects of changes of 'timbre', it was proposed to compare the performance of 'musical' and non-musical' groups in certain pitch matching situations, to find whether it was possible to discriminate between groups by these means.

Finally, a method was devised by which all subjects could attempt matching on a simple instrument, so that the influence of such a 'personalised' means of tone production could be investigated.

Preliminary Considerations.

For the purposes of the following, it is essential that the difference between 'pitch' and 'frequency' be clearly understood, as the two are in no way synonymous. Pitch is the entirely subjective impression of 'highness' or 'lowness' experienced by a person attending to a tone. Frequency is an objective measure of the periodicity of a tone. Although as a general rule we can say that increases in frequency result in the subjective experience of increasing 'highness' in pitch, the relationship is in no way simple. Equal increases in frequency do not result in equal pitch jumps, but in successively smaller and smaller pitch jumps. Other things being constant, a doubling of the frequency results in a pitch jump of one octave. Thus, to move from a tone with the pitch A, the standard frequency of which is 220 cycles per second, to a tone of pitch A, 440 cycles per second, requires a jump of 220 c.p.s. (Herz) To move from A to A , a jump of a further octave, requires an increase in frequency not of a further 220 herz, but of 440 herz, since the standard frequency of A' is 880 herz. Frequency is therefore, from the point of view of the physicist, a logarithmic function of pitch. This relationship of frequency to pitch is liable to considerable distortion in certain circumstances in the testing situation, and it is vital that all such influences be controlled.

The relationship between pitch and frequency is liable to disruption whenever there are changes in intensity (amplitude). The effects of amplitude on subjective pitch have been described by Stevens and Davis, who produced a table of equal-pitch contours showing apparent change in pitch as a function of change of amplitude. (N.B. 'Loudness' is the subjective experience resulting from the amplitude of a tone. A change of amplitude, which is an objective measure, results in a subjective change in 'loudness'. The relationship between loudness and amplitude, and loudness and frequency, will be discussed later.) The table of equal pitch contours is reproduced overleaf.



"Contours showing how pitch changes with intensity. The percentage change in frequency necessary to keep the pitch of a tone constant in the face of a given change in intensity can be taken as a measure of the effect of intensity upon pitch. Pitch in this case is the parameter, as indicated by the numbers attached to the curves. The ordinate scale was arbitrarily chosen so that a contour with a positive slope shows that pitch increases with intensity". (After Stevens, Introductory

Acoustics, Van Norstrand, N.Y., 1933)

The graph is intended to show that "for low tones, the pitch decreases with intensity, but, for high tones, the pitch increases with intensity". (The left hand axis on the graph is, in the opinion of the writer, incorrect, and taken at face value makes the graph uninterpretable. In a later article, in the Journal Acous. Soc. Amer., 1935, Stevens presents basically the same graph but with a different vertical axis, which is more meaningful. Basically, the axis should show that for the lowest intensity levels there is very little distortion of apparent frequency (pitch), but that distortion is progressively greater as intensity increases. On the right hand axis, the writer has added a different set of figures, based on Stevens later article, which it is believed makes the graph understandable.) Basically, then, an increase of intensity (amplitude) makes high notes sound higher, and low notes sound lower. There is also an interaction between signal intensity and subjects judgements about the duration of auditory stimuli. (Tanner, Patton and Atkinson, 1966)

In the testing situation it is important that the interaction of amplitude and apparent frequency (pitch) be controlled. (Cohen, A., 1961) Similar effects are observed in the relationship of loudness to amplitude. (Stevens and Davis; Fletcher, H. 1935) Stevens and Davis write of this function, "In general, the lower the frequency the more rapidly does loudness grow as a function of intensity, at least for intensities below the 100 db level. Thus a tenfold (20-db) increase in the intensity of tones whose loudness is 0.1 sone produces, in a 50-cycle tone, a two-hundredfold increase in loudness, but only an eleven-fold increase in a 1000-cycle tone". There is a similar type of relationship between loudness and frequency. Fletcher and Munsen (1937) produced a graph of equal-loudness contours, for tones of various loudness levels (in phons), as a function of frequency. In general terms, it can be said that with amplitude held constant, very low or very high tones sound subjectively less loud than tones in the middle of the frequency range. From the graphs it would appear that the greatest loudness level is experienced with tones of from about 500 c.p.s. to about 5000 c.p.s. There is fairly rapid fall off in loudness below and above this range. The apparent attenuation outside this range is greater for louder tones; for tones 10 to 15 d.p. above threshold there is an almost linear relationship. A more general coverage of the phenomena associated with the subjective perception of pitch is given by Ward, W. D. (1954).

Aims of the Experiments.

There were three main aims. These were, firstly, to show the effects of differential timbre on various pitch matching tasks; secondly to demonstrate what differences, if any, were apparent in pitch matching experiments involving a variable pitch whistle as opposed to the more orthodox function generator, (audio-oscillator); and thirdly to show what differences, if any, existed between musical and non-musical groups, in the performance of the above tasks.

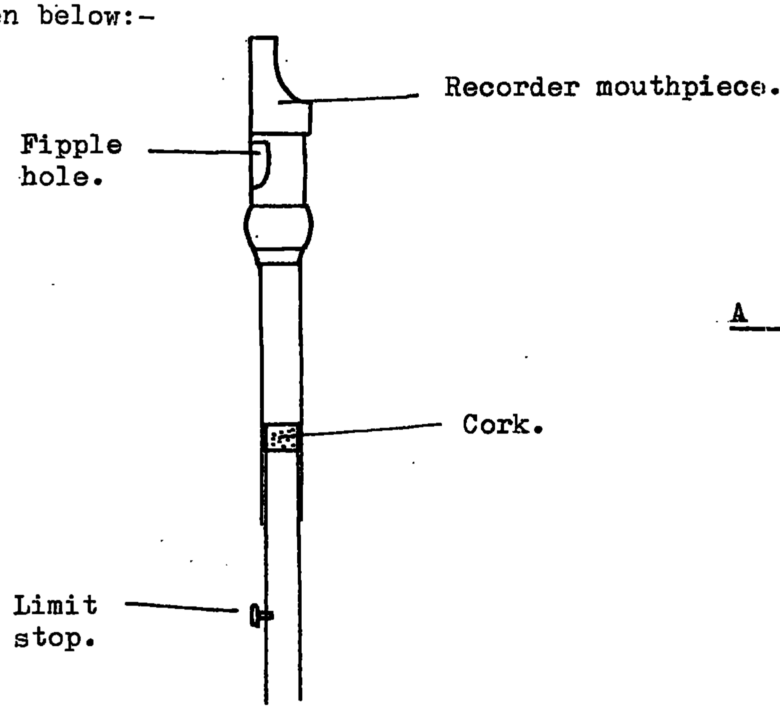
Pilot Study.

Method:

A pilot study was carried out initially, which though unsatisfactory in certain respects, suggested ways of carrying out the main experiments. The pilot study was basically an attempt to relate ability to match tones from an audio oscillator, with either a variable pitch whistle, or by singing, to various factors of home environment as measured on an ad hoc questionnaire. Details of the equipment used in the pilot study are given below:-

Variable pitch whistle.

This was constructed from two close fitting lengths of brass tube, which formed a variable length pipe, and a cheap, plastic recorder purchased from a branch of Woolworths. Basically, construction consisted of cutting the brass tube to a suitable length, (to give the required "lowest" note), and inserting this by means of an adaptor into the sawn off mouthpiece of the recorder. A cork was inserted into the top end of the inner tube to produce a stopped pipe. Finally, a screw acted as a limit-stop at the top end of the range, since it was found that pushing the inner tube in to its fullest extent produced undesirable harmonics, especially on overblowing. A description of the operation of stopped and open pipes can be found in The Sensations of Tone, Helmholtz, 1877 (2nd English Ed. 1954). The resulting 'instrument' is similar to the 'Swannee Whistle', devised many years ago and now difficult to obtain. A diagram is given below:-

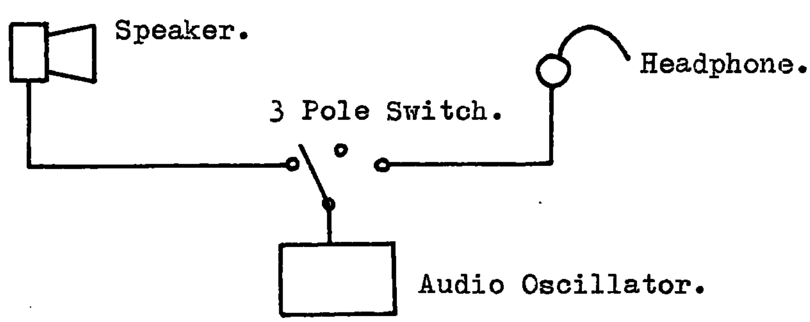


A Variable-Pitch Whistle.

The subject blows gently into the mouthpiece, and can alter the pitch of the note by sliding the inner tube in (to increase frequency) or out (to decrease frequency). The frequency range of the whistle was from about 500 c.p.s. up to about 1200 c.p.s. There were very slight variations with individual performers, causing variations in this range of no more than a semitone; the range was over one octave, from approximately B up to D^2.

Electronic Apparatus.

In the pilot study(6) this was very simple, consisting of an Advance Type J-2 audio-oscillator, a single telephone type headphone, and a 10" Stentorian double-cone loudspeaker in a suitable cabinet. There was a simple switching mechanism. A diagram is given below:-



Apparatus for Pilot Pitch-Matching Experiment.

Using the above apparatus it was proposed to relate pitch matching (on a variable pitch whistle, and vocally), to standard sine-wave tones from an audio-oscillator, to scores on a questionnaire on musical background.

Procedure.

The experiment was conducted in a normal office type room. Background noise registered 30 to 35 d.b. during periods of no signal (registered on a Dawe battery operated transitor sound-level meter, type 1400E; fast response, 'normal' weighting). The subject and the experimenter sat on opposite sides of a table. The dial of the oscillator faced the experimenter and the speaker faced the subject.

Initially, the subject was allowed a few moments to accustom himself to the sound, and to the mode of operation, of the whistle. Experimental procedure was then as follows: A stimulus tone was presented for ten seconds. (Timing was by a stopwatch.) Subjects had been instructed, upon cessation of the tone, to attempt to match the tone with a tone of the same pitch by using the whistle. This procedure was repeated nine times. In the nine trials, three stimulus tones were used, presented in random order. Following this a further nine trials were given which necessitated the subject matching by singing tones. For the 'vocal' part of the test, the subject was initially asked to sing or hum (whichever he found easier) a tone which he could 'comfortably' produce; the three test tones were then given about the tone produced, and were simple 'doh, me, soh' intervals, using the initially sung tone as the third.

Subjects.

These were undergraduate students, from a variety of disciplines. Two groups were used:

(1) An unselected group of 20, comprising eight females and 12 males.

(2) A group selected for its 'musicality', using membership of the music department, or a high degree of executive skill in classical performance as criteria. This group comprised 10 subjects, 6 male and 4 female.

All subjects were given the identical testing procedure; following which they were asked to complete a questionnaire.

Questionnaire.

This was concerned with the musical background of the subjects, and involved questions on musical behaviour, concert attendance, preferences, past training, and certain specifically musical questions concerning time signatures and key signatures. The questionnaire was very lengthy. An arbitrary scoring system was used.

Measuring Technique.

Subjects responses on the pitch matching task were measured in an unsatisfactory way, though the achieved accuracy was later shown to be quite high. The subject produced tones on the whistle which were attempts to duplicate standards of 600, 750, and 950 c.p.s. These tones were chosen as being well within the range of the whistle, yet leaving room for considerable error. They were not harmonically related. The experimenter measured the frequency of the 'sung' or 'whistled' sounds by means of the earphone. With the switch in a left hand position it was possible to hear the subject's response in the uncovered (non-earphone) ear, and the oscillator in the other, via the earphone. The subject's tone was 'tracked' by E on the oscillator, and matched by beats. On the 'singing' section, octave errors were scored correct; on the whistle, octave errors were possible only on the 600 c.p.s. tone. (These were very rare in all cases.) The main criticisms here are the uncontrolled background noise, lack of control of amplitude by E in making the measurements, lack of control over the subject's distance from the speaker, and the fact that errors were not quantified but judged merely 'right' or 'wrong'.

Results.

It is not proposed here to give any detailed account of the findings, since more detailed studies are described later; and also the experimental design was not entirely satisfactory, as indicated above. The main findings were as follows:-

(1) When matching on both the whistle, and with the voice, the musical group scored significantly more hits than the unselected group.

(2) No relationship was found between questionnaire and matching scores for the unselected group.

(3) No relationship was found between questionnaire and matching scores for the musical group.

(4) When both groups were analysed together, there were significant differences between questionnaire scores for the two groups. Since the questionnaire confounded factors of formal learning, musical habits, and environmental conditions, this finding cannot be interpreted in any meaningful way.

(5) Those subjects with high scores on the whistle scored high scores on singing also. Ability to match by singing, however, did not necessarily imply the ability to match on the whistle. This is probably because subjects were more familiar with the timbre of their own voices than with the timbre of the whistle; and certain subjects were unable to isolate the variable of pitch from a novel sound stimulus which differed in timbre and frequency from the standard tone.

Conclusions.

In the two pitch masking tasks, it appeared that music students were superior as a group to a group of unselected students, though individuals in the unselected group obtained high scores. It also appeared that the change from whistle to voice was more easily made by the music students than by the others, though the small numbers make this only a tentative conclusion. It is interesting that no relationship was found between matching scores and questionnaire scores for the unselected group, but this may have been due to the over-complex nature of the questionnaire. On this point, follow-up studies were performed by Davies and Jackson (1967, unpublished report), in which musical attitudes and social environment were examined, and Davies and Billings (1968, unpublished report), in which musical behaviour and the presence of 'circumstances favourable to musical development' were examined. Neither of these minor studies showed any relationship between the musical and environmental variables. Both studies involved administration of questionnaires to samples of 50, and 150, college students respectively. The reason why no relationship was found is almost certainly a function of the sample used. By selecting college students, we are in effect selecting subjects from a very restricted social bracket. Bergel (1962); Kahl (1953) and Mulligan (1952) all show that a high degree of restriction exists in the way in which different social classes are represented in universities and colleges. A study by Davies (unpublished dissertation, 1966) showed a relationship between musical behaviour and scores on a standardised test of musicality, however.

The findings from the above pilot study suggest that more detailed comparison of musical and non-musical groups in various pitch matching situations might shed light on differences in the ways in which the two groups deal with pitch material. Also, the need for more rigid controls over the stimulus stiuation is apparent if conclusions are to be anything more than tentative.

The failure of the present questionnaire, and the need for deriving standards of reliability and validity coupled to the computation of accurate loadings, (which are necessary if any questionnaire is to be used with confidence) led to the abandonment of any further questionnaire studies in this context.

First Pitch Experiment.

This follows on directly from the pilot study, and is an attempt to obtain meaningful, quantitative data from pitch matching situations, Many workers, (e.g. Békésy, Ritsma, Stevens, Guttman, et al.) have performed matching experiments of various kinds. A general finding is that complex tones are more easily matched than sine wave (pure) tones. The relative ease of matching complex sounds is attributed to the greater harmonic richness, the upper partials being a guide to matching which is not present (except for aural harmonics:- see Aural Harmonics and Combination Tones, Stevens and Davis, Hearing, p. 184) when pure tones are used. Ritsma, R. J. (1966) and Schouten, J. F., Ritsma, R. J. and Lopes Cardozo, B. (1962) showed that pitch matchings between a periodic complex sound and a pure tone are difficult. It was hoped to investigate further the effects of matching tones of similar, or of different complexity; and also to find out if the responses of musicians differed in any way from the responses of an unselected sample.

The wider aim of these experiments, namely, to provide a basis for the construction of a group testing battery in 'classroom' conditions, should be born in mind. This aim was a primary consideration in the choice of the type of auditory experiments to be performed. No threshold tests of the type performed by Stevens (1947), von Békésy (1960), Guttman (1962, 1963, 1965) et.al., were performed here. The reasons why peripheral threshold measurements are inapt in musical testing have been discussed earlier. Also, no tests of perfect pitch, in any form, were included. For discussion of perfect pitch and its various forms, see Neu, D. M. (1947), Bachem, A. (1948), Boggs, L. P. (1907), Mull, H. K. (1925) (this study showed that people could be trained in the acquisition of perfect pitch), Weddell, C. H. (1941). Though in the past, the possession of perfect pitch was regarded as being important in the development of musical skills, this is not the modern view. A sense of <u>relative</u> pitch is regarded as being more important. Although a great many musicians, especially those with early musical training, seem to possess perfect pitch of one form or another, (Bachem (1940)), and exceptionally musical children possess absolute pitch in a majority of cases; (Teplov (1966)), it is by no means indispensible in the development of an ear for music (Teplov). Shuter also concludes that "the possession of absolute pitch is not a necessary component of a high degree of musical talent". Music teachers of the very highest calibre (Dame Ida Carroll, Principal, Northern School of Music; Ronald Wright, Senior Clarinet Tutor, Northern School of Music; Sidney Fell, Professor of Clarinet, Royal College of Music, Manchester; Edgar Hunt, Trinity College of Music), in personal communications, have stressed the importance of relative pitch as opposed to perfect or absolute pitch. Perfect (absolute) pitch is thus regarded as being outside the scope of this study.

Method.

It was proposed to measure the pitch matching abilities of musical and non-musical groups. Pitch matching would be to standard tones which were pure; the matching being carried out under two conditions, namely with either the variable pitch whistle described in the pilot study, or with an audio oscillator. By using headphones and a sound-proof cubicle, interference was to be minimised. The experimental layout, and apparatus used, are described below:-

Variable Pitch Whistle.

As in the pilot study.

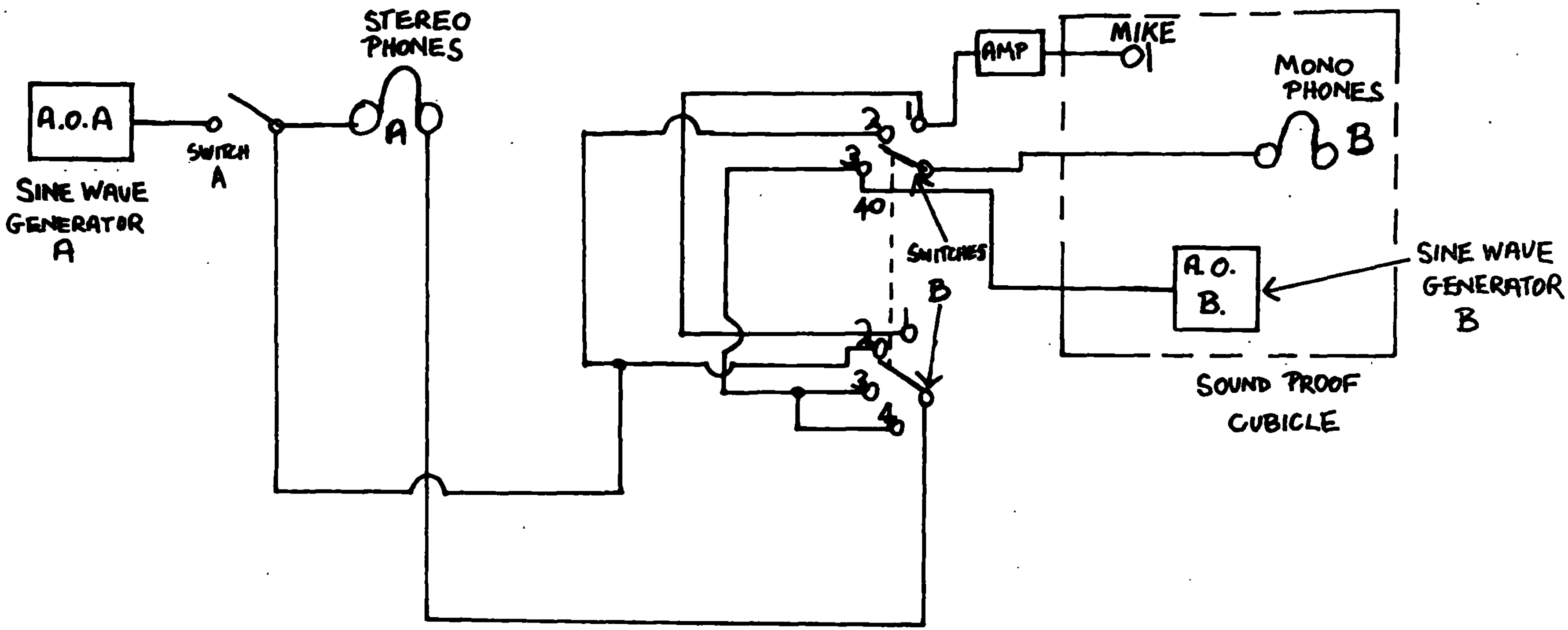
Electronic Apparatus.

A general diagram of the layout is given overleaf in fig. 2 . The layout enables the experimenter (E) to select one of five possible conditions. With switches B in position no. 1, and switch A in the closed position, the subject hears the sound he makes (in both ears) on blowing the variable-pitch whistle (v.p.w.) close to the microphone; whilst E hears dichotically the sound produced by the subject (S) on the v.p.w. and the tone from the audio oscillater (a-o) A. With switches B in position 2, and switch A closed, both S and E hear, in both ears, the tone from a-o A. With switch A closed, and switches B in position 3, S hears the tone from a-o B in both ears, whilst E hears dichotically tones from a-o A and a-o B. With switches B in position 4, and switch A closed, S hears nothing, whilst E hears tones from a-o A and a-o B dichotically. Finally, with switch A in the open position, no sound is heard by either S or E.

Oscillator A was an Advance Type J-2; (described in the pilot study); oscillator B was a Marconi Audio-Tester, Type TF 894A. Both generate pure (sine-wave) tones.

The sound-proof cubicle was an I.A.C. (Industrial Acoustics Company) sound-proof chamber, type 401 A3.

The stereo-phones A, used by the experimenter, were Koss stereophones, model Pro 4. The 'phones used by the subject



EXPERIMENTAL LAYOUT FOR FIRST PITCH MATCHING EXPERIMENT.

fig 2

(mono-phones B) were Sansui stereophones, type SS1, with identical inputs to each ear.

The amplifier was a battery driven, transistorised Sinclair, model Z12 (total harmonic distortion .01, on 9 volt supply, 50 ohms load). All switches were brought together on a simple panel which could be operated by E. The layout permits no verbal communication from E to S. Since trials in which S had to match using the whistle or the oscillator were rotated in a random fashion, a subsidiary circuit was installed, not shown on fig. 2. This was used to operate a simple light display situated in front of the subject, which indicated whether a 'whistle' response or an 'oscillator' response was required.

The microphone was of the ribbon type, by Reslosound, model RBT/L, 30 - 50 ohms; a floor stand was used.

Procedure.

Subjects were introduced into the sound-proof chamber (S.P.C) and allowed time for acclimatisation. A trial period was given with the v.p.w. Subjects were then given instructions; a written version of the instructions appeared in the s.p.c. to help subjects remember the sequence of events. This sequence was as follows:- The subject was seated comfortably, and fitted with the mono-phones in a correct manner. Upon commencement of the first test item, the subject heard a tone, in both ears, of ten seconds duration. Upon cessation of the tone, the subject had to match a tone, from either the whistle or a.o. B, as directed, to the stimulus given, in terms of pitch. After the subject had achieved a setting on either the whistle or the a.o., he was asked to signal that he had made his final choice on that item. When using the whistle, he did this by saying "O.K." into the microphone; when using the a.o. he depressed the 'range change' switch once, which had the effect of introducing a high pitched note into the tone finally selected. Subjects were asked to hold the note finally selected for a few seconds before signalling that final choice had been made, and then to wait for a visual signal, given by E via a mirror, before returning either the whistle or the a.o. to a standard position. The dial on a.o. B. was covered by a circle of card, so that the subject would receive no cues from the position of the pointer.

Six randomly arranged trials were given on both the v.p.i. and a.o. B. making a total of twelve trials per subject. The standard tones, against which the matching took place, were sine wave tones from a.o. A., having frequencies of 950, 750, and 600 c.p.s. The range of the v.p.i. was a major consideration again, in the choice of standard frequencies.

A problem encountered in this situation is the difficulty of controlling, and measuring, the sound pressure level that actually impinges on the subject. A measure of the output to the earphones is not likely to be of any use, unless corrections are made for the response characteristics of the earphone. Earphones can be calibrated, but the process is costly and time consuming. Also, different heads produce different degrees of acoustic impedance, so that the values obtained from calibration may not be an accurate guide to the sound level which impinges on the subject's ear when the headphones are worn. To overcome this problem, calibration can be performed using a box which has the same acoustic properties as an average human head. In some headphones, provision is made for the installation of a calibrated microphone, which gives a measure of sound pressure level inside one of the headphones, whilst being worn. None of these methods was available to the experimenter. Subjects were instructed, upon commencement of testing, to set the level on a.o. B to a 'comfortable level'. The tone used for this setting was 750 c.p.s. When S had done this, E matched the same tone from a.o. A. to the level set on a.o. B. by dichotic listening. When two tones of the same frequency are heard dichotically, an imbalance in the sensation level is reflected in an apparent change from the median in the location of the sound source. For a tone of 800 c.p.s., 40 d.b. above threshold, a difference of about 2 - 3 d.b. is sufficient to cause a noticeable displacement from the median position of the sound source. (Stephens and Davis, 1936). Once these settings had been achieved, no further change in the output of either oscillator was made. Bearing in mind the range of frequencies used, and the fact that no extremes of amplitude were used, from Stephen's and Davis' data, very little distortion of pitch due to interaction with loudness would be expected. From the data presented on page 131 , it can be observed that in the present situation, assuming that no extremes of amplitude were selected, pitch and frequency have an almost linear relationship. Similarly, data by Fletcher and Munson (1933) shows that, for the tones used here, any change in loudness as a result of change of frequency is likely to be of the order of only one or two decibels for the 600 c.p.s. tone, and virtually no change for the others. Finally, it must be added that ultimately it was hoped to use results obtained for the derivation of material for use in a group test battery. Clearly, any tests which proved sensitive to the differences mentioned above would be unlikely to prove satisfactory in a classroom.

Once S was seated in the s.p.c., and the preliminaries were completed, a signal given by E (via a mirror) indicated that the first trial was about to be given. E then selected the appropriate frequency on a.o. A, with switch A open, and switches B in position 2; closing switch A then presents the stimulus tone from a.o. A.

to both S and E. After ten seconds, timed by stopwatch, switches B were immediately thrown to position one (for a whistle response) or position three (for oscillator response). Simultaneously, the light display, indicating either 'whistle' or 'oscillator' was activated. S would then try to reproduce by the appropriate means a tone of the same frequency as the stimulus. Since there is a virtually linear relationship at the frequencies and amplitudes used, frequency is an accurate indicator of pitch in the present situation. No time limit was imposed, and the subject was allowed to continue his matching attempts until he gave the signal indicating that the last tone played was the one to be judged by E. S then waited a short time (six or seven seconds) before returning either the whistle or the oscillator to the starting point. Marks on the cardboard screen attached to a.o. B. indicated the beginning and end of a frequency range within which all tones lay. This range was the same as the range of the whistle. The 'zero' or starting point for each trial was taken as the lowest point in this range, a tone of about 500 c.p.s. A groove was filed on the tubes of the whistle to indicate the correct 'zeroing' point.

When the subject indicated that he had made his final choice on a particular matching trial, the frequency chosen was measured by E, using a dichotic listening procedure. With switches B in position 1 (whistle response) or position 4 (oscillator response) E hears the sound produced or selected by S, and the standard tone to be matched, from a.o. A, dichotically. Any difference between the frequency of the stimulus and the response tone is readily discernable. The error is found by E performing a pitch match of his own. The tone from a.o. A is matched to the tone selected by S; but note that E hears the tones simultaneously whereas S only hears one tone under all conditions. Matching is performed by E, using the method of beats. This method proved to be extremely accurate, and far more rapid than using an electronic counter. After E's matching, the error (difference between tone selected by S and the standard) was read from the dial of a.o.A.

As a check to the accuracy of this method of measuring the error, a series of matchings, over a far wider range than used in the study, (to check for the presence of octave errors) was performed by E. After each matching, the reading given by E was checked, using a Racal Universal Counter Timer, Type SA 535. Results of this check are given below. The table gives the difference between the measurement using the 'method of beats' and the frequency measured by the counter timer, on ten trials; stimulus (matching) tones were selected randomly by an assistant.

Trial	Error (c.p.s.)
Trial 1.	Error 0.4 c.p.s.
2.	0.1
3.	0.1
4.	0.3
5.	0.0
6.	0.2
7.	0.0
8.	0.2
9.	0.1
10.	0.2

Mean error in dichotic matching by E = 0.16 c.p.s. This level of accuracy is regarded as being very satisfactory.

As a further check on the overall pitch matching accuracy of E, two further series of ten trials were run, in which E performed the task normally performed by S. (i.e. in the sound-proof cubicle, with matching carried out after the cessation of the stimulus tone; that is, using the remembered pitch of the stimulus, and not the method of beats as above). It was anticipated that errors here would be greater than for the dichotic listening procedure. Results are given below; tones were again selected randomly by an assistant.

Trial	Error	Trial	Error
Trial 1.	Error 5 c.p.s.	Trial 1.	Error 2 c.p.s.
2.	1	2.	1
3.	4	3.	3
4.	3	4.	3
5.	15	5.	5
6.	12	6.	8
7.	5	7.	8
8.	1	8.	9
9.	12	9.	2
10.	0	10.	6
Mean error 5.8 c.p.s.		Mean error 4.7 c.p.s.	

The left-hand series involved matching sine-wave to sine-wave tones; results on the right were obtained in matching square-wave tones to sine-wave tones. Measurement was again made using the counter-timer.

The calibration of a.o. A. was such that a degree of interpolation was necessary on all readings; measurements were therefore taken to the nearest 5 c.p.s. on the experimental runs.

Sample.

This comprised university students of both sexes, from different disciplines, and in various stages of advancement in their courses. Two groups were chosen; one group was selected as a 'musical' group, using membership of the music school or a high degree of performing skill, as criterion; the other group was a randomly selected sample.

Results.

In the randomly selected group, 56 subjects were measured in performance on the whistle and on the oscillator. For each subject errors were placed in two categories. These were, errors when matching with the whistle, and errors when matching with the oscillator. Total deviation (error) was computed, between chosen tone and stimulus tone, ignoring sign, for each subject. Mean error (in c.p.s.) per subject was also calculated.

Similar procedure was adopted for a sample of seventeen 'musical' subjects. Total and average error in c.p.s. was calculated for whistle and oscillator for each group. These are given below:-

		Total error.	Mean error.	N.
Musical group	Whistle	629 c.p.s.	37 c.p.s.	17
	Oscillator	289 c.p.s.	17 c.p.s.	17
Non-Musical group.	Whistle	43305 c.p.s.	773 c.p.s.	56
	Oscillator	20295 c.p.s.	362 c.p.s.	56

Differences between the two groups (musical and non-musical) in terms of matching performance on both whistle and oscillator, are extremely significant.

If we adopt a nul hypothesis that there is no difference between the likelihood of errors on the whistle exceeding errors on the oscillator, and oscillator errors exceeding whistle errors, we would have a chance expectation that 50% of subjects would perform better on the whistle, and 50% would perform better on the oscillator. We can compare the expected ratio with the obtained ratio and see if the difference is significant. The results of Z-tests comparing the ratios are given below:-

	Subjects with more errors on whistle.	Subjects with more errors on oscillator.	Z-score.
Musicians.	9	8	0 (not sig.)
Non-Musicians.	49	7	6.26 (very highly sig.)

The above is a more meaningful way of looking at the results, since t-tests between errors in c.p.s. for the two groups result in degrees of significance that are almost meaningless. It is clear from the above that non-musical subjects on the whole find the whistle much harder to match than the oscillator; whilst the musical group appears to show little difference between the two. (However, two of the musicians did find the whistle very much harder, which accounts for why errors in c.p.s. are different for oscillator and whistle in the musical group.) Also, the base rate of errors in c.p.s. is much greater for the non-musical than for the musical group. Two musicians did find the whistle more difficult. If results from these two are taken out, mean error on the whistle for the musical group drops from 37 c.p.s. to 11 c.p.s. Certain individual musicians then do display the same difficulties as those experienced by the non-musical group.

Discussion.

The above results show that matching on the whistle was more difficult than matching with the oscillator. The difference in the degree of difficulty, however, was greater for a non-selected sample than for a sample of musicians. The difference in the difficulty of matching under the two conditions could be due to two factors, however. Firstly, the tone of the whistle was complex whilst the oscillator tone was pure. Unlike tones are harder to match than like tones (Ritsma, 1966). Secondly, the whistle involved a degree of psychomotor activity analogous to normal musical performance; whilst it could be argued that the oscillator (involving just the turning of a knob) did not involve this to the same degree. Thus, whilst the experimental task discriminated well between a musical and a non-musical group it is not possible to say what factor in the situation was responsible for this. The second pitch matching experiment clears the situation to some extent.

In a sense, the results obtained above are paradoxical. Pitch matching tasks involving complex to complex matches have been found to be more easily performed than sine to sine matches. The relative ease with which two complex, as opposed to two sine-wave, tones are matched has been attributed to the greater harmonic richness of the complex tones, which is said to give more cues to the subject. (See references on page 143) In the present situation, the subject hears a stimulus tone which is pure; this is presumably difficult to 'locate' in terms of pitch, due to the absence of upper partials. He then has to perform a match using either sine-wave or complex tones. In the latter case, he should be able to locate the pitch of the note better than in the former.

However, due to the difficulty of locating the pure stimulus tone initially, one would not necessarily expect pitch matching on the whistle to be better than on the oscillator, even taking into account the presence of upper partials. It is very difficult, however, to explain why pitch matching on the whistle should in fact be worse. The above findings, and those of Ritsma et al., make sense if one assumes that upper partials only influence pitch matchings to any great extent when the wave form of the compared tones is similar in both cases. Where wave forms are perceptibly different, the difficulty of comparing in one dimension (pitch) sounds which differ in two dimensions (pitch and timbre) swamps any effect of timbre in an absolute sense.

To clarify the situation it is necessary to standardise the psychomotor aspects of the task, by devising means of synthesising complex tones electronically; and also to discover whether the results obtained above are specific to sine stimulus tones and complex matching tones, by alternating trials of this type with trials having complex stimulus and sine matching tones. The second experiment provides evidence on both these points.

Second Pitch Matching Experiment.

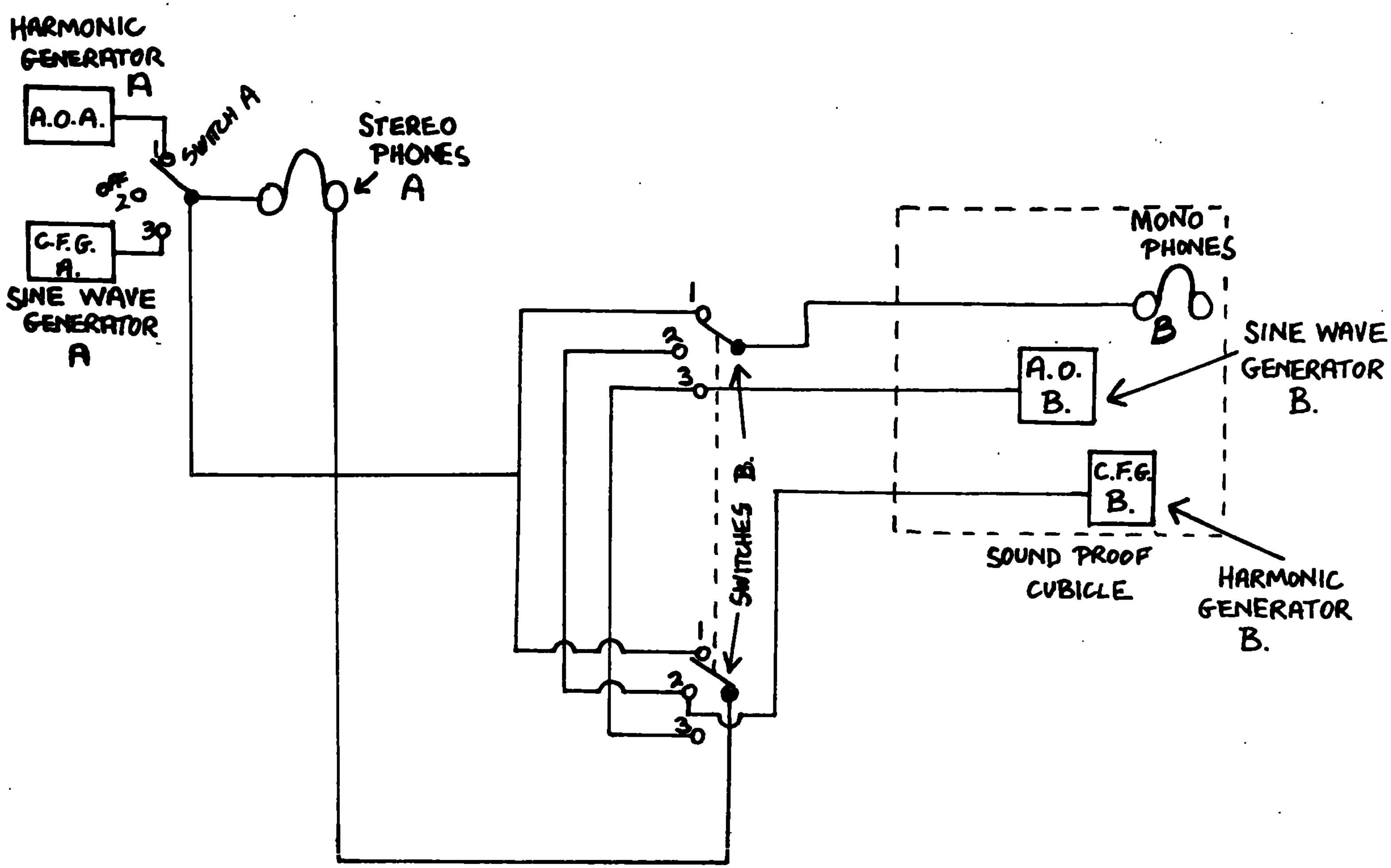
The basic aim of this experiment was to throw further light on the influence of timbre in pitch matching situations. Problems arising from the comparison of matchings on the variable-pitch whistle and matchings on the audio-oscillator, in the first pitch matching experiment, have already been outlined.

Method.

This was broadly the same as for the first pitch-matching experiment, except that the variable pitch whistle was replaced by a specially constructed harmonic generator. Experimental layout, and electronic equipment is described below:-

Experimental layout.

A diagram giving the general layout is included overleaf in fig. 3. In this layout, E can select seven possible combinations of events. The complex function generator A (c.f.g.A.) and the sine-wave generator (s.w.g.A.) can be selected by switch A, and presented to S's mono-phones. In this position, E and S both hear the stimulus in both ears. For presentation of stimulus (from either c.f.g.A. or s.w.g.A.) switches B must thrown to position one. After presentation, switches B are thrown to position



EXPERIMENTAL LAYOUT FOR SECOND PITCH MATCHING EXPERIMENT.

fig 3

two for a c.f.g.B. response, or position three for a s.w.g.B. response. When measuring the error of the subject's matched tone, the same dichotic listening procedure used in the first pitch matching experiment was employed. Although E has a choice of either c.f.g.A. or s.w.g.A. as a 'tracking' tone, the pure tone was used in all cases, as pure tones produce stronger beats with the tone to be tracked. With switch A in position two, no tones are heard by either E or S. The layout permits matchings of the following type:-

	Stimulus Tone.	Response Tone.
1.	Sine wave	Sine wave.
2.	Complex	Complex.
3.	Sine wave	Complex.
4.	Complex	Sine wave.

Production of Stimulus Tones, and Response Tones.

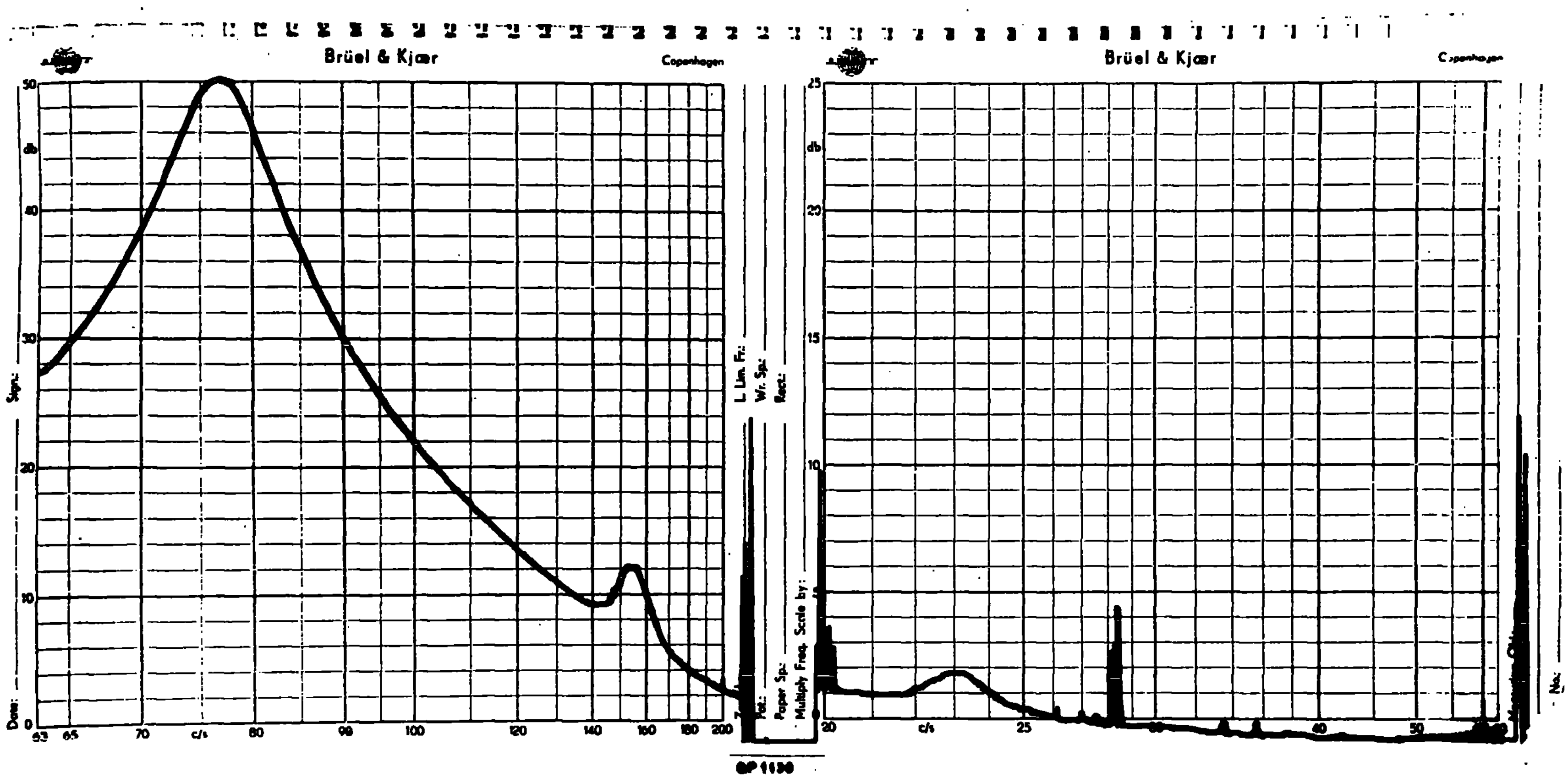
The sine wave tones were produced on commercially available signal generators. These were of the following type: s.w.g.A. was an Advance Type J-2, as used in the first study; s.w.g.B. was the Marconi Audio-Tester, also used in the first study.

Production of complex tones presented a serious problem, and considerable time was spent in finding a solution. It was felt that for future experiments, a layout which had flexibility both in frequency and in harmonic content was desirable. To this end, a pair of generators, one having continuous frequency but discrete harmonic capability, and the other having discrete frequency steps but continuous harmonic capability, was thought desirable.

An all solid state van Gogh low (audio) frequency function generator, type TV 1A was used to produce sine, square or saw-tooth, wave forms over a continuous audio-frequency range. This was an easily portable machine, operated from six 1.5 volt bar batteries.

The other complex function generator (harmonic generator) was built in the workshops, and gave five different operating frequencies, with variable intensities on four partials. A harmonic generator giving three fundamental frequencies, and controllable intensity over five partials is briefly described by Hickman, Music Education Research Papers, number 3, Electronic Apparatus for Music Research. The unit used in the present study gives a frequency range slightly greater than that used by Hickman, but one harmonic partial fewer. A circuit diagram of the unit in final form is given in fig. 4. Note that no diagram is given here of the power supply unit which was entirely conventional; also, the switches controlling frequency are shown with three poles only, for the sake of clarity. There is no reason for limiting the number of poles to three, and in the present generator five were used. Considerable time was spent in the development of the final version of the generator. Earlier attempts to synthesise tones using sonic wheel, (production of tones and partials from small generators, adapted from electric motors; these were linked by a system of gears and driven by a single motor capable of maintaining a steady speed. The result was a particularly dismal howling sound) and beat-frequency oscillator (which failed due to instability) were unsuccessful.

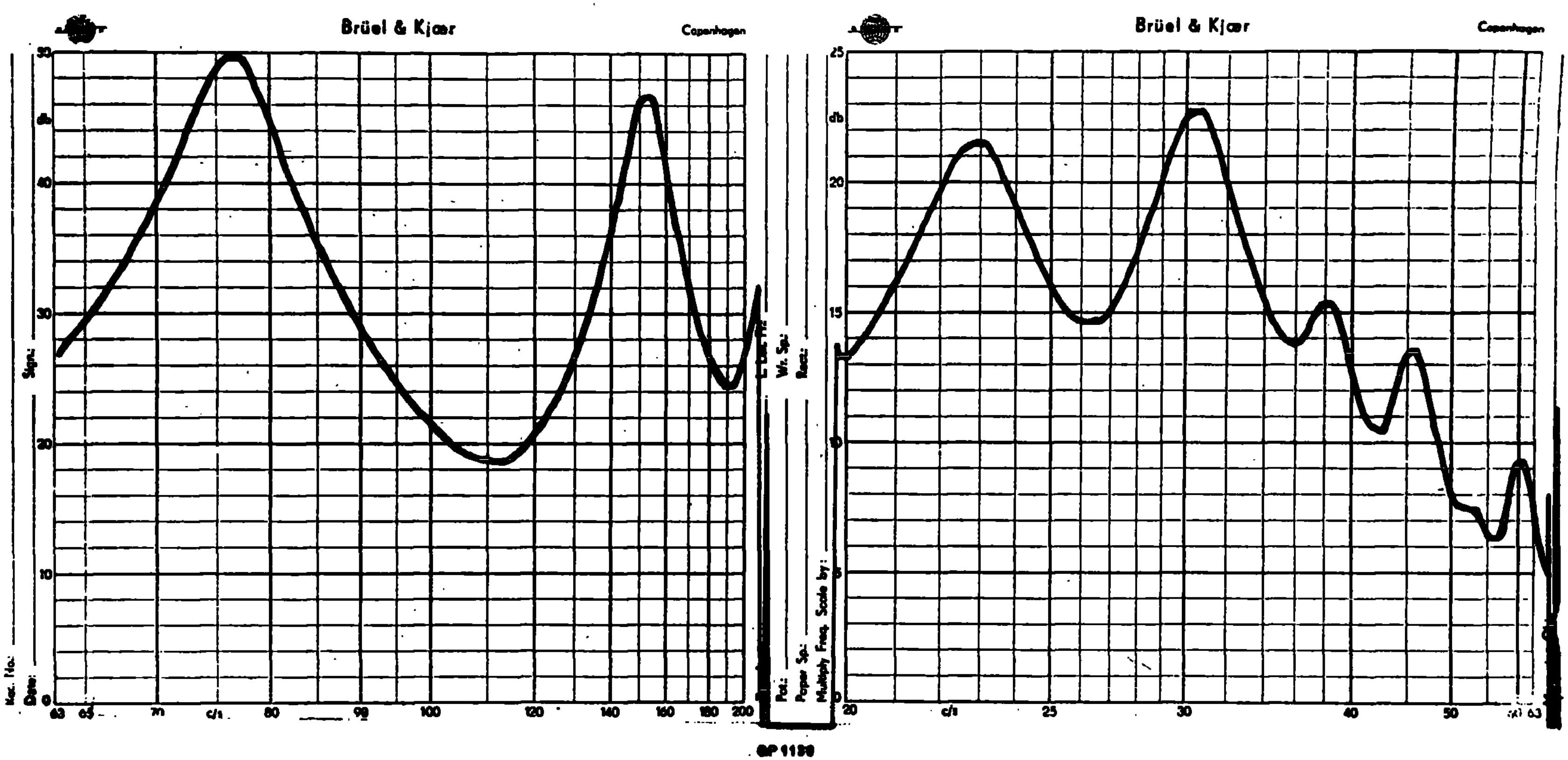
The generator used consisted of an inductance-capacity oscillator, and three multipliers. The first upper partial was obtained from a doubler, the second from a tripler, and the third from another doubler linked to the first doubler. High Q coils were used throughout. A tuned filter was placed between the output and the oscillator. Coils were wound on a home-made coil winder, and also by means of a technique involving a lathe. Total cost to build was about £15. The performance of the unit, though probably not good enough for extremely accurate measurement of thresholds or similar psychophysical measurements, was entirely satisfactory for present purposes, where a wide range from subjectively pure to complex tones was required. Graphs are given in fig.5 & 6, showing the performance of the unit when producing (1) its 'purest' tone, and (2) its 'most complex' tone. The graphs were produced on a Bruel and Kjoer wave form analyser, by direct line from the harmonic generator. Improvements could be made by adding further resonant filters between stages.



HARMONIC GENERATOR IN 'PUREST' MODE.

ALL UPPER PARTIALS SET TO ZERO.

fig 5



HARMONIC GENERATOR IN 'MOST COMPLEX' MODE.

ALL UPPER PARTIALS SET TO FULL.

fig 6

Procedure.

Subjects were introduced into the sound-proof chamber and given time for acclimatisation. Subjects were then given their instructions. A card bearing the more important points, particularly the need to 'zero' the apparatus after each trial, was displayed in the chamber. Experimental procedure was much the same as in the first pitch experiment. Subjects were presented with a tone for ten seconds. This tone could be either sine wave, from s.w.g.A or complex, from the harmonic generator in the 'most complex' mode. Upon cessation of the stimulus tone, subjects hear a tone from either s.w.g.B. (sine) or from c.f.g.B. (complex), depending on whether E throws switches B to position 2 or 3. He has to match the new tone to the pitch of the stimulus tone. The accuracy of S's judgements is measured by the same dichotic listening procedure as used previously.

The tones used were this time dictated by the frequencies which the harmonic generator was capable of producing. The five frequencies used were 650, 700, 750, 850, and 1050 cycles per second. Trials on the four conditions given on page 160 were alternated randomly. Each combination of frequency and timbre was presented three times, giving a total of 60 trials. The experiment was carried out with each subject tested on two successive days, half the items being given in each session. Two sessions of twenty to twenty-five minutes were thought preferable to a single lengthy session.

Subjects.

These were 20 randomly selected persons from the university, including students of both sexes, and laboratory technicians. There were 8 females and 12 males. Examination of the data shows no significant sex difference.

Results.

The main finding is that the results obtained in experiment I, in which matchings on the whistle were found considerably harder to perform than matchings of sine tones from an audio-oscillator, by an unselected (random) group of subjects, are confirmed. Matchings of pairs of tones that are very different in timbre are harder than matchings of like tones; so the hypothesis stated at the end of experiment I, that differences might be due to the different nature of the whistle task as opposed to the oscillator task is rejected. It seems that the main influence was the timbre difference, and not the difference in the physical nature of the tasks.

In the present study, a significant difference was found between matchings of tones that were alike and the matching of tones that were of different timbre. No significant difference was found between the sine-complex and the complex/sine match however;

so the main influence again appears to be timbre difference, with no effect accruing from using sine stimulus/complex matching tones as opposed to complex stimulus/sine matching tones. A table of the results obtained is given below:-

	Both Pure.	Both Complex.	Pure/ Complex.	Complex/ Pure.
Mean error. (in c.p.s.)	118.9	156.7	696	606.9
S.D.	67.79	59.7	310.4	284.1
	'Alike' Matchings.		'Different' Matchings.	
Mean error.	137.45		651.45	
S.D.	66.55		300.9	

The above gives means and standard deviations for all four groups, and also for the two groups of 'alike' matchings and the two groups of 'different' matchings. A table of the significance of 't' is given below:-

Significance of the difference between the means for various groups.

Difference between 'alike' matchings and 'different' matchings.	t = 10.5487	Highly Significant.
Difference between 'both pure' and 'both complex' matchings.	t = 1.836 ($.1 > p > .05$)	
Difference between 'complex/pure' and 'pure /complex' matchings.	t = 0.95	Not Significant.
Difference between 'both complex' and 'pure/complex' matchings.	t = 7.65	Highly Significant.

Whilst the general finding that tones of different complexity are harder to match than similar tones is confirmed, the paradox concerning ease of matching complex tones mentioned earlier is again present in this experiment. The results show that in the present situation, the complex to complex match was performed with greater difficulty than the sine to sine match. The conclusions reached by other workers (Ritsma et al.) suggests that the opposite situation ought to obtain. The result obtained here is therefore of particular interest, though the significance level of the difference between the 'both pure' and the 'both complex' groups is such as to prevent any very definite conclusions being drawn. Nonetheless, it would have been expected that any difference between these groups would have been in the opposite direction from the one obtained.

Discussion, and conclusions from the pitch matching experiments.

During examination of existing tests and measures of musical ability, in an earlier section, the point was raised that certain batteries used stimulus tones in which the variable of timbre was not controlled. The two pitch matching experiments show clearly that the judgements of subjects are very much affected by the nature of the sound stimulus used. The usual argument against this is that the uncontrolled timbre, in the group testing situation, makes no difference, since whatever 'tone colour' is used and however this might vary between items, the stimuli are the same for all subjects. The first pitch matching experiment shows that this is not necessarily so, since a sample of 'high achievement' musicians showed less distortion in their judgements due to timbre change, than did a sample of 'non-selected' subjects. At present, we cannot assume that this difference between musical and non-musical groups is due to some constitutional or hereditary factor in the musical group. Certainly, the musical group has had far greater experience of a wide range of musical tones with different timbres, than has the non-selected group; and evidence has already been presented showing that familiarity with tones is an important factor in determining the degree to which pitch tests of various kinds can be performed. (Stumpf) The implications from this experiment for the construction of test material would seem to be fairly clear; namely, that the use of stimulus material of varying tonal complexity could produce misleading results in certain circumstances. The proposed test battery uses only stimulus material of a sine wave or 'pure' type. The material could have been standardised on any particular wave form; but Teplov showed that the subjective change in the 'size' or 'sharpness' of a tone (low tones tend to be described by subjects as being 'fat' or 'heavy', whilst high tones are described as sounding in some way 'small' or 'thin') is greater for complex tones than for pure tones. This effect is attributed to the differing numbers of upper partials that are audible at different frequencies, resulting in the perception of differing percentages of the total output by the listener. With sine wave tones, provided they are within the normal range of hearing, 100% of the energy is heard.

All items in the proposed test are composed of sine wave material, with the exception of the rhythm test. The further influences of the pitch experiments on the construction of the pitch test for use in the battery will be discussed later.

First Version of the Test Battery.

The Pitch Test.

Existing pitch tests, almost without exception, involve subjects in hearing two temporally spaced stimulus tones, and requiring them to make either a same/different judgement or an up/down/same judgement. The simple detection of change in the above manner is a fairly rudimentary process, which does not necessarily have implications for more complex, and possibly more important, pitch tasks. Basically, the argument is that there are so many cues to a simple change of pitch (changes in subjective quality, loudness, and pitch) that a great many people can perform the task who would be unable to perform a more complex, and from a musical point of view, more pertinent, task such as pitch matching. The capacity to tell that two sounds are different does not imply the capacity to adjust them till they are the same. This latter task is much more important in instrumental performance where playing 'in tune' is important. This point is supported by a considerable body of experimental evidence. The evidence takes the form of studies of the correlations between ability to perform 'same/different' tasks and the performance of other musical tasks. Highsmith (1929), McCarthy (1930), Brown (1928), Mosher (circa 1930), and Lamp and Keys (1935) all performed studies of the above type. Individual correlations range from 0.17 to .49. Average correlation is about 0.335. A detailed examination of these, and other studies, is given in Teplov, La Sensibilite a la Hauteur du Son. Teplov refers to the same/different test situation as being one in which 'la sensibilité differentielle à la hauteur' is involved. His conclusion is that 'nous pouvons dire que la sensibilité differentielle a la hauteur n'a aucun lien avec le sens musical en general, et notamment avec l'oreille musicale'. (Translation: We can say that sensitivity to the difference in the pitch of notes bears no relationship to general musical ability, and especially not to 'the musical ear'.)

For the present test battery, some way was needed of getting away from the standard group testing techniques of pitch testing. The experiments in the sound-proof room had shown that the method used for individuals could discriminate well between a musical group and a non-musical group. However, certain subjects from the non-musical group, without any formal musical training, performed the tasks very well; whilst two low scoring 'musical' subjects showed that formal training was not a guarantee of a high level of performance on the tasks used. There was considerable variability in both groups. The case, that tests in the sound-proof room discriminated between groups on the basis of environmental or experiential factors, has not been shown to be obviously false; but sufficient indications exist to make deeper examination of these methods seem profitable.

Examination of the manner in which subjects performed the pitch matching experiments, described previously, suggested ways in which the method might be adapted to a group testing procedure. The experimental task involved the following:-

(1) The subject hears a tone for several seconds.

(2) Upon cessation of the tone, S tries to keep a tonal image of the note played.

(3) He manipulates an oscillator, or a whistle in such a way that he hears a rising glide tone.

(4) He continues the upward sweep until a 'clang' (in a Helmholtzian sense) occurs between some particular frequency in the glide tone and the remembered tonal image of the stimulus. In the most successful subjects, there is probably an increasing sense of the nearness of the 'clang' at different points on the glide tone.

(5) Since considerable experimental evidence exists to show that 'same/different' judgements on the whole are fairly easily-performed, a failure to score a hit in the above situation will be due to an inability to store the tonal image in most cases, rather than an inability to perceive the 'clang'. The task therefore involves the ability to select from a continuous frequency range, a tone which will coincide with a particular tonal image. This is the type of task performed habitually by performing musicians.

The experimental procedure was adapted to suit the needs of a group test in the following manner. First, subjects would be presented with a stimulus tone, in the same way as in the experimental situation. They would be expected to remember this tone. After a short pause, the subjects would be presented with a glide tone, and expected to locate the stimulus tone at some point in the glide tone. The main problem lay in devising a means by which a subject could indicate whereabouts in the glide tone the 'clang' occurred. In the first version of the pitch test, this was done by producing a long glide tone, which was broken into four segments by short bursts of white noise. Subjects had to say whether the stimulus occurred in the first, second, third, or fourth segment of the glide tone. The duration of the stimulus tones was five seconds; each stimulus was presented twice, with two second intervals between. After the second presentation, three seconds elapsed before the onset of the glide tone. The duration of the stimulus tone was chosen on the basis of papers by Wayne A. Wickelgren (1966), Turnbull, W. M. (1944), Aaronsen, D. (1967) and Bachem A. (1954). Wickelgren shows that although the most accurate perception of pitch takes place in the space of about one second, consolidation of the memory trace takes considerably longer. It appears that, up to 8 seconds duration, the longer the standard tone is present, the stronger the memory trace for that tone becomes. It was thought that two presentations of five seconds, giving a total time of ten seconds, would give time for consolidation without the stimulus tones appearing to be unnecessarily long.

Three versions of the glide tones were used. These tones all traversed the same frequency range, but took different lengths of time to do so. Before hand, it was difficult to see how the different sweep speeds would affect the difficulty of items; thus, although the slowest sweep speed would allow the longest time for comparisons to take place, and also produce the most prolonged 'clang', the time between stimulus and 'clang' would be greater than for the faster sweep tones, permitting greater time for the decay of short-term memory trace for pitch and also interjecting the greatest amount of tonal interference (noise) between stimulus and response. Results from the first test run suggested that the middle speed was most suitable. The frequency range traversed by the glide tones was from 500 c.p.s. to 1300 c.p.s., with .1 sec. of white noise interjected at 500 (start), 677, 888, 1100 and 1300 c.p.s. (stop)

Construction of test tapes.

The frequencies at which interjection took place, and other peculiarities in the material, will be more clearly understood if a brief description of the manner in which the test tapes were constructed is given. All items for all parts of the test battery, except the rhythm test, were constructed from sine wave recordings, using a Revox full track tape recorder, at 15" per second. Ampex professional tape was used throughout. In order to construct the tapes, a tape 'bank' was constructed, consisting of lengths of tape, each bearing a recording of a particular sound used in the tests.

Test tapes were then constructed by splicing together sections of tape in the required order. Whilst this is a very time consuming process, results are extremely worthwhile, since extremely accurate control over length of tones is possible by cutting tape to exact lengths; and also any distortion or wobble due to turning on or switching off oscillators, or stopping and starting tape recorders, is absolutely eliminated.

The glide tones were produced by fitting electric motors (high quality Crouzet clock motors) to the frequency tuning spindle of the Marconi audio-oscillator. Three different motors were used, having speeds of 4, 8, and 12 revolutions per minute. After recording, the glide tone tapes were cut at the correct 'starting' and 'stopping' points, and then divided into lengths of exactly one quarter. $1\frac{1}{2}$" of white noise was then spliced between each section ($1\frac{1}{2}$" = .1 sec. at 15" per second). This procedure ensured that the 'blips' of white noise were exactly equally spaced in time. The frequencies at which the blips occurred were then measured, and are given above. Recording level was kept constant at 50% for all recording.

In the above form the test was administered to five graduate students. As yet, there were no standardised instructions, so the task was explained verbally. It rapidly became apparent that the test was far too difficult; before any trial runs were made in schools, the test was reconstructed so that the glide tone had only three segments, separated by four blips of white noise. The number of possible alternative answers was maintained at four by introducing a 'none' category for stimulus tones that did not occur in any of the glide tones. Six 'correct' responses occurred in each answer category, giving a total of twenty-four items. Within each segment, stimulus tones were spaced at equal distances apart, in a temporal sense. It was thought that those tones occurring in the middle of any glide tone segment would be 'easier' than tones located at the ends of glide tone segments. "None' answers were spaced three above the glide tone range, and three below; extremely high and extremely low tones were contained to give a proportion of easy items. Frequencies used will be given for later versions of the test only, since many of the ones used in this first version were never administered to test groups.

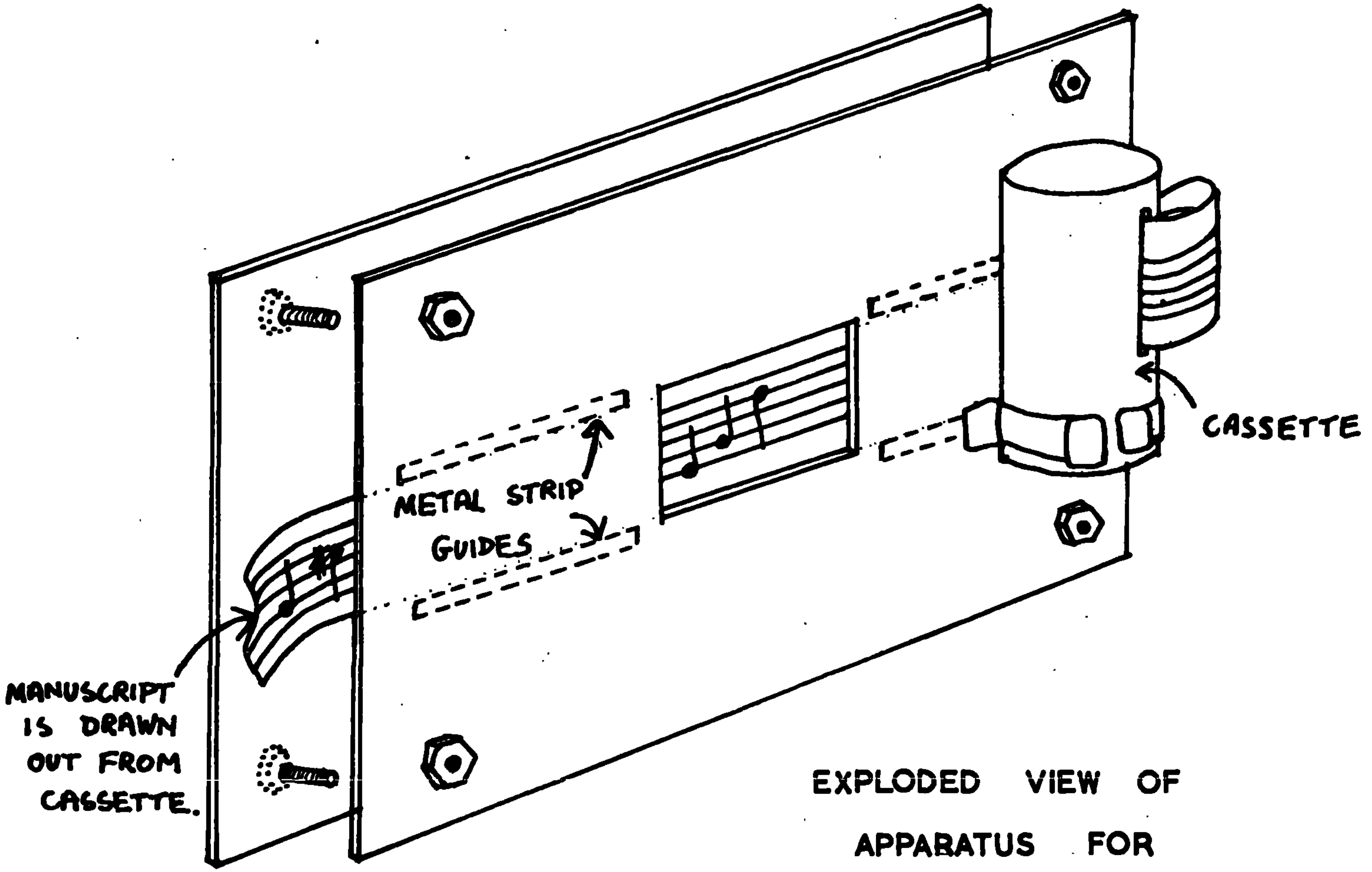
<u>The Tonal Sequence Test (Tonal Memory)</u>.

Existing tests of memory are usually directed toward measuring a persons ability to remember certain characteristics of tonal sequences. For the purposes of this argument, all tonal sequences will be referred to as 'tunes'. The point has been made in section 1, that the often used procedure of presenting two identical or almost identical tunes, and asking subjects to state whether they are the same, or if different to locate an altered note, are in

some ways simply pitch tests with a high ratio of noise to signal. (see pages 24 - 25) A means was therefore sought to test subjects' capacity for recognising whole 'tunes', using a short-term memory method. From the literature on short term memory, there would appear to be a paucity of data on memory for tonal sequences, though a considerable amount of data exists for short term memory for single tones or sounds. (D. J. Corcoran and D. L. Weening, 1967, Turnbull, W. M., 1944, Bachem, 1954, et al.) Quastler however describes experiments involving pianists, in which subjects were presented with musical material of differing degrees of 'organisation'. The material was composed from various tonal 'alphabets' containing from three up to sixty-five notes. Speed was also varied. The number of errors made by pianists increased as a function of speed, and also as a function of alphabet size. (Quastler, H., 1956) Also relevant to the present problem is a study by Miller and Selfridge, 1953, and a study by D. Howes and C. E. Osgood, 1954 and 1954. ?
In the Miller and Selfridge experiments, the effects of varying the degree of statistical approximation to normal English on short term memory for linguistic material was examined. The percentage of the words correctly recalled was found to be a function of the length of the word lists used and the degree of statistical determination in any word list. The method used for producing the statistical language was adapted in the present situation to permit the production of strings of 'statistical music'.

Using 'quasi-musical' tunes of a statistical nature seemed an attractive approach for several reasons. Firstly, it is of intrinsic interest to find out whether the findings of Miller and Selfridge with regard to length and codability would be duplicated for musical memory. Secondly, the method of producing statistical music offered an attractive alternative procedure for producing material with a controlled degree of codability; many workers have been content to make up material themselves using some completely internal criterion, or to draw material from the formal compositions of celebrated composers. Thirdly, the method offered a possible way of controlling the order of difficulty of items.

"Statistical Approximations to Music" were constructed in the following manner. Small groups of high grade musicians, all of whom met some composing criterion, were drawn together. As a further test, a check was made to ensure that they were in fact familiar with musical notation, by asking them to name certain well-known tunes from glancing at a score. Musicians from a variety of backgrounds were used, including jazz players (modern and traditional) and folk musicians, as well as classical musicians. They were seated in a comfortable room, and presented with the piece of apparatus shown overleaf in diagram 7 . Two such contraptions were constructed in order to speed the process. The apparatus, briefly, consists of a cassette which is loaded with several feet of paper



EXPLODED VIEW OF APPARATUS FOR PRODUCING "STATISTICAL MUSIC."

fig 7

bearing the five lines of the musical stave. The cassette is mounted on a 'sandwich' of two pieces of hardboard, one of which has a viewing window. The stave paper can be drawn past the window by the user. In operation, a person is presented with the apparatus, and, according to the level of contextual restraint, sees several notes in the viewing window. He writes one more note that he thinks might reasonably be expected to follow those displayed and then pulls the paper along to cover up one note. The apparatus is then passed on to the second member of the group, and so on.

Subjects were told that all notes were assumed to be in treble clef, and that all notes must be of standard crotchet duration. Apart from these two points, the subjects had a completely free hand in choosing their notes; they were simply instructed to add notes as though they were trying to construct "the strongest melodic line". 1st, 3rd, 5th and 7th orders of approximation were constructed. In the third order material, subjects had to add notes to groups of three that were presented to them; in fifth order, they were presented with five notes, and so on. A list of the material derived by the above method is given in appendix 2.

The above proved a satisfactory method of deriving musical material of different degrees of codability, but with no particular stylistic bias. Length could easily be varied by cutting the material into suitable chunks. The major remaining problem of how to devise a test of true tonal memory rather than a pitch test with high noise to signal ratio, (see p. 24 - 25) was approached in the following manner. The major requirement of any satisfactory test of tonal (melodic) memory is that it must test memory for whole tonal sequences, and not put all the emphasis on one isolated element of a tonal sequence. It can be convincingly argued that considerable deviation of individual tonal elements can be tolerated by tonal sequences without the sequence as a whole losing its identity. Many writers (Shuter, Teplov, Wing, Bentley) give instances of the attempts of children to sing short tunes, songs, nursery rhymes, etc., stressing the point that tonality is often imperfect. The fact that an individual tone is wrong, however, does not justify the assertion that the child cannot remember the tune. If such an assertion were true, no-one could have recognised the tune in the first place. The inclusion of wrong tonal elements remains a problem of interval perception, and not of melodic memory; though naturally, as the number of wrong tonal elements increases, the effect on the integrity of the 'courbe melodique' also increases. The following test was devised as a means of removing the overwhelming emphasis on individual tonal elements found in certain memory tests.

Firstly, subjects would hear a short extract from the store of statistical music. After a short pause, a longer piece of the quasi-music would be played. Subjects would then be asked to indicate whether or not the second extract contained the first extract. In this way the emphasis is placed on entire tonal sequences, and individual elements become relatively unimportant.

In the earliest version of this test, tunes of five different lengths were chosen, comprising 8, 10, 12, 14 and 16 notes. The tunes were randomly picked from the store of statistical music. The shorter tunes, which subjects were asked to recognise in the longer sequences, comprised half the number of tones contained in the long sequences. For each length, one item each was taken from each order of statistical approximation, making a total of twenty items. In this earliest version, subjects were simply asked to answer 'yes' or 'no', giving a guessing score of 50%. Each stimulus tune was given two presentations.

As with the pitch test, the earliest version was administered to a small group of graduate students. Scores were very closely grouped about 50% correct, suggesting that (a) items were on the whole far too difficult, and (b) some way of decreasing the chance score was needed. It was also apparent that the test was far too long to be of practical use with young children. The following modifications were made before the test was used in schools. Firstly, the number of items was cut to fifteen. The length of the individual items was cut radically to make the task easier. Five items with stimulus tune lengths of three notes and longer tune lengths of six notes, and ten items with respective tune lengths of four and eight notes, were included. Finally, in order to reduce the guessing score, a new element was introduced into the tests; subjects were asked to count 'how many notes were left over after the shorter tune had stopped' on every occasion when they answered 'yes'. At this stage, it was by no means certain whether children would understand such a complicated task. Below is a diagramatic representation of the task to be performed, whenever the correct answer is 'yes':

Short Tune.	Long Tune.	Correct Answer.
[B G D]	F [B G D] C' A	[Yes | 2] (Number of notes left over)

Statistical approximations were distributed randomly and in equal numbers over the fifteen items, except that first order approximations, which should have been the most difficult, occurred only three times instead of four. The answers were distributed in the proportion of eight 'no's' to seven 'yes's'. Number of notes left over on the 'yes' items varied from one to three.

Again, test tapes were constructed by splicing together prerecorded segments of tape.

The Interval Test.

In the present battery, the test of interval can be regarded as replacing the test of harmony or chord analysis which frequently occurs in existing test batteries. The evidence for leaving out tests of harmony or chord analysis has been given at length in section two. The importance and the function of interval measures have been discussed in the same section. Unlike the tests of tonal memory and of pitch memory already described, there are very few precedents for the type of test material described here. Interval measures have been used by Madison (1942) and by Hickman (1969), but both these workers used intervals in a harmonic (simultaneous presentation) context, rather than in a melodic (consecutive presentation) context. (Full description and discussion of these methods appears in section two). The test proposed here uses the melodic approach to interval; the nature of the test material suggests that a better, more self explanatory name for items of this type would be 'auditory transposition'. The Lundin Musical Ability Tests are the only other group of tests that include items with a directly transpositional bias. (Lundin, 1949) In the Lundin tests of Melodic Transposition, subjects hear pairs of melodies. The first melody is in a different key from the second melody, and one or more notes in the melodic line may be changed in the second playing. Subjects must say whether the two renditions of the tune would be the same or different if they were both in the same key. This task is extremely complicated, and involves two quite separate abilities. In essence, the task is exactly the same as the type of memory test pioneered by Wing and Seashore, but involves a transpositional element as well. It will be remembered that the Wing test of memory bears a close resemblance to a pitch test having a high noise to signal ratio. The Lundin test certainly overcomes this objection, but raises another in its place; one can describe this type of test simply as a test of interval (in a melodic sense) with the same high noise to signal ratio.

Below is given a diagram showing the situation that obtains in the Lundin transpositional test. The diagram should be compared with that on page 24 which pertains to the Wing memory test.

Tune 1. Key 1.	A–C 3rd	C –G dim 5th	G–B 3rd	B–C 7th	C–F 4th	F–G 2nd	G–D 5th	D –F 6th
Tune 2. Key 2.	B–E 3rd	E –A dim 5th	A–C 3rd	C –E 6th	E–G 3rd	G–A 2nd	A–E 5th	E –G 6th
Same or Dif.	S	S	S	D	D	S	S	S

The above is a diagramatic representation of the Lundin type of test, in which the two compared tunes are viewed as a series of intervals following one after the other. Note that the alteration in the fourth cell of tune two automatically causes a change in cell five. In order to arrive at a correct answer of 'different' the subject must compare cells four and five of tune two with a memory copy of cells four and five in tune one. To detect the change, he must perceive that in some way the interval of a seventh in tune one is different from an interval of a sixth in a different key in tune two. The correct performance of this task means that the 'signal' has been picked up by the subject; and the remaining material is merely 'noise'. Note however that in order to perform the task correctly, the subject must be able to make accurate interval judgements, and must also be able to hold the complete tonal sequence in short term memory. Thus, two quite separate abilities are involved, and the absence of either one makes performance of the task impossible. A paper by J. P. McKee and D. A. Riley (1962) suggested that a transposition test might be suitable for use with young children. Children with a mean age of 6.7 years were trained to discriminate a 'correct' stimulus tone from an 'incorrect' tone. When the actual pitches of the two tones were altered, but the tonal relationship between them was kept constant, 50% of the children showed 'spontaneous transposition'. A similar effect was observed for pairs of tones which differed in amplitude.

In the present battery, a test of tonal memory is already included. The proposed interval test is based directly on the type of problem set in cells four, and cells five, above. Briefly, the proposed test would involve subjects in making judgements about certain melodically presented intervals. However, it is vital that any proposed problem can be solved without the need for any knowledge of interval names, or of any other musical symbols. This is achieved by devising a task which places emphasis on the apparent 'distance' between two tones of different pitch. An analogy may be made in a different context, with visual experiments in which subjects have to judge differences in the distance of objects (Holway and Boring, 1941, 1942) or in the apparent size of perceived objects. (Hastorf, 1950; Bruner and Goodman, 1947). Note, that apart from the general comparative method used in these studies, no serious parallel with the present material is intended.

The first version of the interval test was as follows. Subjects were presented with a stimulus interval. After a short pause, further intervals were presented, each having a different root note from any other, and also different from the root of the stimulus. One of the non-stimulus intervals would be the same as the stimulus interval, but in a different key. The subject performs the task by listening to the stimulus interval and remembering how far apart the pitches of the constituent tones are. He then has to pick out the subsequent interval which contains tones with the same tonal separation. Each item consisted of a stimulus interval, of two consecutively produced tones. The second tone was always higher than the first, so that all intervals were ascending. The stimulus interval was presented twice, with three seconds between presentations. After this, three 'comparison' intervals were presented, with two-second spacing. These intervals were also ascending. Subjects had to indicate which of the three comparison intervals had the same tonal separation as the stimulus interval, by answering one, two or three.

A full chromatic scale from tonic to tonic contains thirteen tones. Each possible combination of tonic and one other note from the octave above was used twice in the stimulus tones. This made a total of twenty-six items for this test. It was impossible to pair each stimulus with every other possible combination of intervals without the length of the test reaching astronomical proportions, (22464 trials). Comparison intervals were therefore assigned randomly to the stimulus intervals.

In this form the test was administered to a small group of graduate students, prior to being tested in any schools. This task proved by far the most difficult of the tests devised. Average score barely exceeded the chance score of 8 - 9. The test was altered in the following manner. Firstly, the third alternative 'comparison' tone was removed from all items, and possible answers reshaped into the form 'one', 'two', or 'neither'. The eleven items which had seemed to be the most difficult were discarded, leaving fifteen items, with five 'right' items in each response category. A certain amount of re-organisation was necessary to achieve this state of affairs.

The same splicing technique was again used in construction of the test tapes.

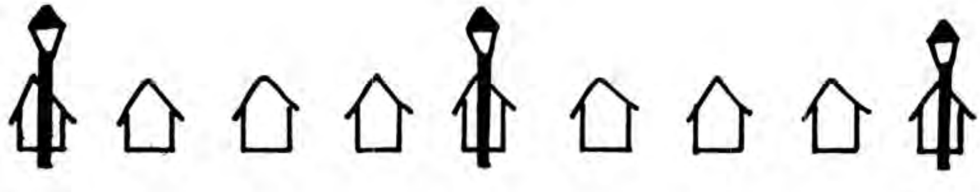
The Rhythm Test.

This test was the most difficult to produce, since it is almost impossible to isolate any rhythmic aspect which is culture free. Thackray's approach of breaking 'rhythm' down into simple component parts, much as other writers have done with tonal abilities, is clearly the best approach. In the present battery, however, a single useful measure is needed. A major problem is that the classical music of the Western world is, with a few notable exceptions, extremely elementary from a rhythmic point of view when compared with African, Afro-Cuban, Moorish, and Indo-Asian rhythms. In a study by C. S. Myers (1904) an assessment was made of the performance of a rhythmic task by a Sarawak Malay. The study was carried out 'in situ' in Borneo. Myers noted that a variety of gong-like instruments were used by the Malays in their music making. One large gong, the tawak, was beaten in what appeared to be a more or less random fashion, regardless of who was playing it. In one study, Myers substituted a morse-key for the tawak, and then analysed the time intervals between strokes. Though at first, he was inclined to think that performers on the tawak had poor rhythmic sensibilities, he rapidly changed this opinion and his final conclusions were extremely perceptive. He concluded that the musicians he heard were capable of remembering, using, and improvising upon, rhythmic structures in which the main pulses were separated by different, and varying, time intervals. (This contrasts sharply with Western music, where the music is almost invariably formed around measures made up of two, three, four, five, or occasionally seven, equally spaced elements.) Myers goes on to write: "This faculty (the one described above) they carry to a degree which lies so far beyond the power of civilized musicians, that the latter may reasonably be sceptical as to the possibility of its occurence among less advanced people". Many other examples of a similar type may be given: Indian music splits its rhythmic sequences up into long lengths containing as many as twenty or thirty beats known as the 'tala'. This may be subdivided into groups of different lengths, such as three, thirteen, seven and four, giving a basic meter of twenty-seven beats; modern jazz is characterised by extreme rhythmic complexity, based on rhythmic styles from all parts of the world, but with a predominantly African/Afro-Cuban influence. Often, the basic meter is completely unstated, and becomes submerged in cross-rhythms which give the untutored listener a feeling that the rhythm is constantly variable, or quite random. As a final example, the writer possesses a tape recording of a young, very musical V.S.O. worker, rehearsing with a choir of Africans. The song being sung falls naturally into two measures of triple time and one of four, the sequence of ten beats being repeated. In his attempts to make the choir sing it in a manner which he believes to be 'right', the V.S.O. worker tries to impress a quadruple meter on the tune, by clapping loudly. The clapping clashes sharply with the native rhythmic accompaniments, and amid shouts and hoots of glee, the rehearsal breaks up. If he had perhaps listened more carefully, the mere concordance of the natives' own rhythmic accompaniments should have convinced him that they knew what was going on, even if he did not. It is understood that he eventually resorted to teaching them hymn tunes.

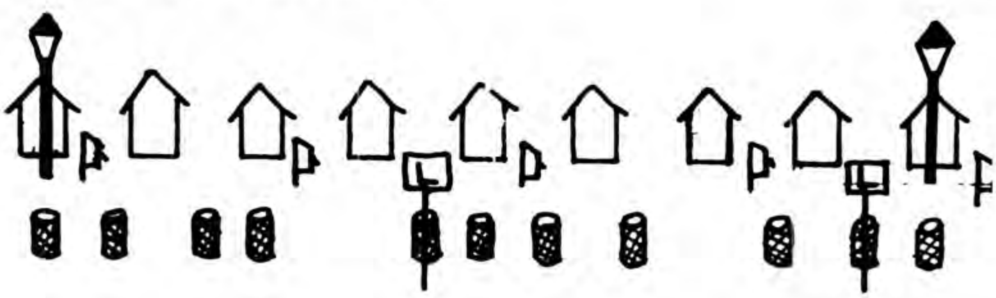
In a single test, it is impossible to measure all aspects of rhythmic ability; the main problem is to find a basic rhythmic ability, which is as culture free as possible, and yet has the greatest possible importance among rhythmic abilities. Examination of existing test material failed to produce any line of approach that might meet these criteria. As a result of a factor analytic study of his own tests, Thackray suggests that the most satisfactory test for measuring general rhythmic ability, where only one test is required, is one "in which the subject is asked to reproduce the rhythm of a melody". Such a test is unsatisfactory in the present context, however. The main attempt in the present work has been to isolate separate components as far as possible; the notion of intermingling rhythmic abilities with tonal memory, expecially when factorial studies suggest that separate factors exist for rhythmic and tonal abilities (see section 1), does not therefore fit in with the present scheme. Though no studies of the interaction of rhythmic with tonal material have been performed, it is possible that in Thackray's proposed test, good melodic memory would certainly be a help rather than a hindrance in the performance of such a 'rhythm' test.

The test proposed here arose from a consideration of the differences between 'metre' and 'rhythm'; and the feeling that a fundamental rhythmic ability, common to the complex tala of Indian music, as well as to the duple, triple and quadruple times of Western music, involved the ability to infer the presence of the 'metre' from complex 'rhythmic' patterns. In this context, metre is intended to describe the basic steady pulse, which underlies a rhythmic structure that may or may not be regular. To understand the nature of the proposed task it is vital that this distinction be understood. An attempt at explanation follows:-

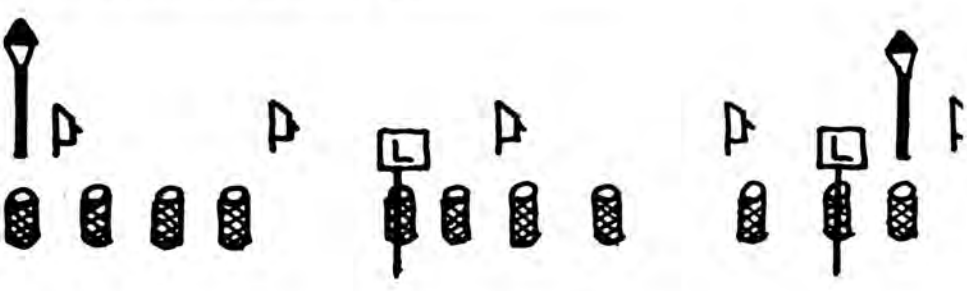
In a certain town, the lamp posts are separated by a certain number of houses. Different streets have different numbers of houses to each lamp. The streets represent different time signatures, or 'metres'. A certain street has a lamp with every fourth house. (Regular accent on first beat)



In this street, but not necessarily in others, every two houses share a mail-box (one AND two, three AND four). There ought to be three litter bins to every two houses, but some are missing. (Stated or implied three across two.) Occasional litter bins are signposted, but not in any regular manner; though the signposts always occur right in front of the bins (purely aesthetic accentuation of certain elements in the three across two). One day a lamp post is removed. (First beat accentuated by omission.) The street now looks as follows:-



Suppose that we now demolish all the houses, and leave behind just the other objects, as shown below:-

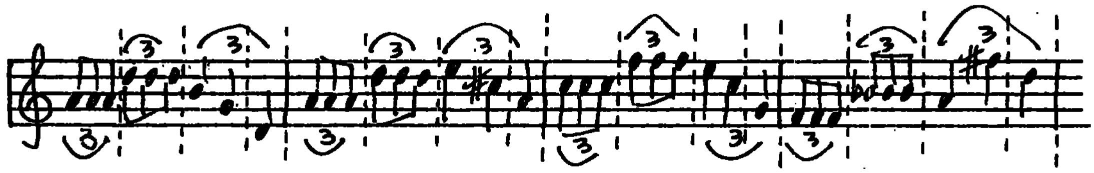


A person familiar with the ways in which lamp posts, mailboxes, etc. can occur will be able to infer from these groupings the basic metre of four houses which underlies them.

The proposed rhythm test presents subjects with simple arrangements of 'mail boxes', 'litter bins', etc. (rhythms) and asks them to infer how many houses occur between 'lamps' (metre). In practice, this takes the form of tasks involving the comparison of simple tapped metres with simple tapped rhythms.

Clearly, in some circumstances, the arrangements of the rhythmic elements are ambiguous, and two possible metres emerge. This is often used to great effect in musical composition, with the result that the perception of the listener switches automatically, and involuntarily, between the two possible ways of hearing the rhythm, in a way almost analogous to an 'auditory' Necker cube. Leonard Bernstein's "I want to be in America" from West Side Story is a well-known example of this. The two ways of hearing this tune are illustrated below, and the vertical dotted lines show the two possible arrangements of 'houses' or pulses of the basic metre.

Implied triple time.



Implied quadruple time.

For the purposes of the present test, it is imperative that no such ambiguity be present, to eliminate confusion. In the first version of the rhythm test, the task was as follows. For each test item, subjects heard three paired presentations of a metre, or regular time pattern with the first beat of each measure accentuated by a bell, and an irregular rhythmic pattern. The metre was the same for each presentation, but the rhythmic pattern was variable. Only one of the rhythmic patterns would fit with the metre. Subjects had to state whether the first, second or third pair was the one that fitted. Altogether there were twenty items, with five items each in duple, triple, quadruple and quintuple time. No 'compound' time signatures were used, as these are all combinations of the above. A theoretical guessing score of five was obtained by introducing a 'none' answer category, for items in which none of the pairs of metres and rhythms fitted.

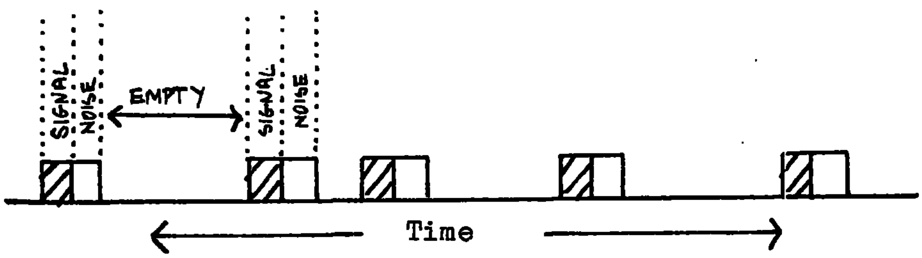
The time signatures, or metres, were recorded from a metronome with a syncronised bell. Six measures (bars) of each rhythm were presented each time. In terms of duration, this meant that the quintuple metre was over twice as long as the duple; however, the subject establishes the metre by the clicks occurring between any two rings of the bell. The bell is the one and only cue to this. By using measures as the criterion, rather than temporal duration, there is the same informational content, and redundancy, in each item regardless of time pattern. Rhythmic patterns were produced by tapping on a wooden block; note that none of these stimuli have any temporal duration (apart from the decay of the bell). Thackray emphasises that any rhythmic pattern is characterised by the relative loudness, the temporal spacing, and the <u>temporal duration</u> of elements. In accordance with these three characteristics, Thackray's tests, and those of other workers (Bentley) present rhythmic material using tones rather than shorter, punctate, stimuli, in order to give each element a certain temporal duration. The present battery adopts a slightly different approach, for the following reasons:-

It should be remembered from the outset that the test proposed here is different in nature from any of the Thackray tests, and places its emphasis on a subject's ability to attend to the <u>position</u> of various rhythmic elements, relative to a certain metre. In such a task, duration of elements is consigned to a relatively unimportant position, when compared with position (temporal spacing) and accentuation of elements. With the emphasis on the <u>onset</u> of elements, it is of little consequence whether the time between elements is filled or empty. If the correct inter-element spacing is removed then amplitude and duration give insufficient cues for the establishment of the correct rhythm; and even in tests using elements of certain durations, there is still a finite pause between elements, in order to indicate where the onset of a new rhythmic element occurs.

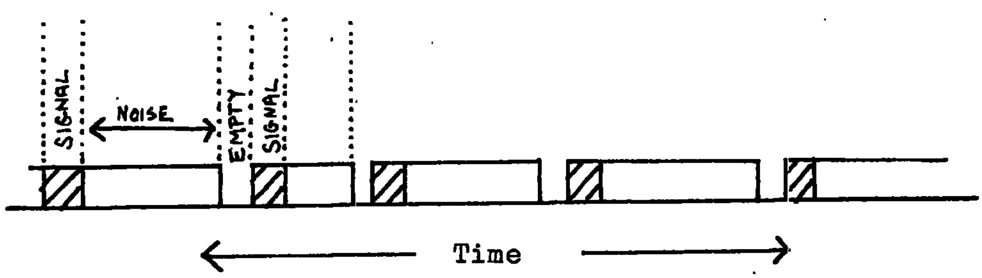
If there were no pause, elements would merge into continuous tones of longer duration, with a certain amount of amplitude modulation. Note also that in existing tests, the elements having duration above and beyond a punctate stimulus, are presented as tones, rather than white noise or a similar non-tonal sound. The interaction of rhythmic and tonal material, for which separate factors may exist, is not known.

In the proposed test, and also in the tests of Bentley (rhythm), the information required by the subject in order to perform the task comes entirely from the onset of rhythmic elements. Therefore, in the present situation, rhythm is taken to imply something that remains unchanged by differences in duration of elements, provided accent and temporal spacing are kept constant. Note that in A.B.R.S.M. examinations, subjects are asked to clap or tap the rhythm of a tune played on the piano. This is a meaningful task, even though the examinee uses punctate stimuli as opposed to the examiner's tune composed of elements of much longer duration; and if the examinee's clapping indicates that he has correctly perceived the onset times of the different elements he is judged to have perceived the rhythm correctly. Given then that we can devise a satisfactory test using the onset times of different rhythmic elements as criterion, there are definite advantages to using punctate rather than continuous stimuli.

It is desirable that in such a testing situation, factors due to attention or accuity be kept minimal. Regardless of the nature of the test stimuli, information about the onset of rhythmic elements is extracted very soon after the onset of each element. Thus even where elements have considerable duration, the information about their onset is obtained, if at all, during the first few milliseconds of presentation. (Guttman used click durations of 70 microseconds. These were readily perceived by subjects. (1965)). Below is a 'signal detection' type diagram, showing the situation obtaining when punctate stimuli are used for presenting the rhythmic material:-



Where non-punctate stimuli, having greater durations, are used, the situation is as below:-



The latter is clearly a much 'noisier' system than the former. The difference between the signal and no-signal condition is much greater for the punctate than for the continuous stimuli, so the former is a clearer, less confusing, mode of presentation.

The rhythm test, in the form described on page 193 was tested on a small sample of graduate students and technicians. It was very clear from the outset, that the test was far too long in its present form. Before preliminary trials were held in schools, the first version of the rhythm test was reshaped as follows. Instead of a separate presentation of the same metre before each comparison rhythm, only one metre presentation of six measures was given at the start of each item. After three seconds delay, the two comparison rhythms were given, separated by two seconds. Subjects had to answer one or two, or 'neither' if neither of the comparison rhythms fitted with the initial metre. The number of items was reduced to fifteen, by reducing the number of times each time pattern occurred from five to four; and the number of trials involving quintuple time, which proved the most difficult, from five to three. Five 'correct' response items occurred in each response category. The speed for all items was ♩ = 144, which is a fairly bright andante.

In constructing the test tapes, a splicing procedure was again used. All parts of the rhythm test were recorded acoustically, whereas items in the previous sub-tests were recorded electronically (direct) to reduce distortion. Metres were recorded from a Maelzel (Swiss) metronome, which possessed a synchronised bell. The early recordings showed that the bell was insufficiently loud; a hand operated, spring loaded (for constant amplitude) bell was substituted, synchronisation being by hand. Rhythmic patterns were tapped by the writer, using lengths of dowel on a hollow wood block, constructed in the workshop. Several minutes of both rhythmic patterns and metre were recorded, and the most accurate renditions selected for inclusion in the test tape. Rhythmic patterns used were selected by the writer, from a variety of sources, including classical, folk, and modern jazz music. Complex ethnic rhythms and alternating or ambiguous rhythms had proved too difficult even for some musical subjects in the earliest versions.

Earliest Test Version:- Preliminary Run.

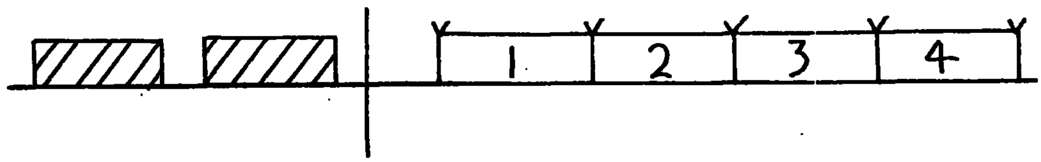
Before any serious attempt was made to administer the test to large samples, it was necessary to try out the test with groups of children in order to find out if the tasks demanded by the tests were intelligible to the children. To this end, a shortened version of the test was administered to children in the Middlesborough area, through arrangement with the Durham University Department of Education. For this pilot run, only the first five items from each sub-test were included. The only aim of this study was to discover if the tasks involved were in fact meaningful to children within the age group for which the battery was intended.

Administration: The Answer Blank.

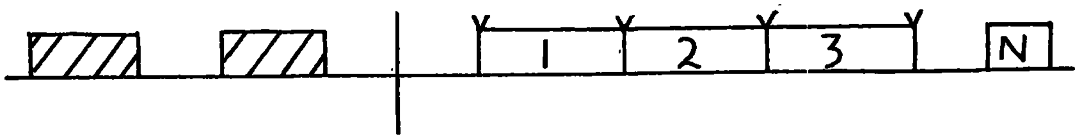
It was thought advisable to produce a new style answer blank for the test battery, which would present more information, in non-verbal form, than existing answer blanks. The Wing answer blank is perhaps not ideally suited for younger subjects. Though instructions are given verbally on the tape, they are also written out on the test sheet. In the present battery, no such redundant information was included. This was also advisable to prevent any reaction on the part of non-readers, or those who only read with difficulty. Similarly, the type of layout used by Wing in the chord analysis test was avoided, since experience with the tests showed that some younger subjects tended to follow the columns across rather than down, or simply to cross out the multiple alternatives in a random fashion. The Bentley asnwer sheet is more satisfactory, since the choosing of multiple alternatives is avoided. In the present blank, however, it was hoped to supply information about each event taking place in any one item, rather than merely supplying a series of squares in which answers are placed.

To do this, the time scale of each item was represented as a horizontal line. Any events taking place in that item (tones, clicks, etc.) are represented by blocks drawn on the line. Answer tones are distinguished from stimulus tones by shading or filling in the stimulus blocks. By following each item from left to right, in a way analogous to reading, the subject sees pictorial representation of each sound he hears. This tells him where he is in each item, helps to prevent him from becoming 'lost', and gives a clear indication of when he has to make his answer.

There is no point in including complete answer blanks for anything but final versions of the test. The diagram below should give a clear indication of the form of the earlier blanks , however. The first illustration shows the pitch test blank in its <u>earliest</u> form. Only one item is shown.

7. 

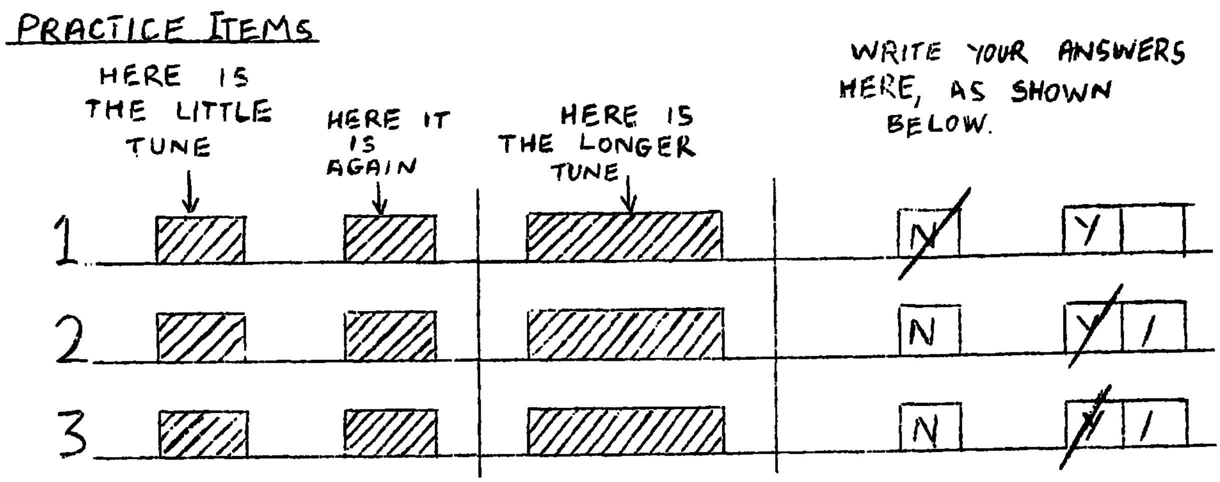
The two shaded blocks represent the two stimulus tones. The vertical stroke separates stimulus from answer tones; the long rectangle represents the glide tone, and the arrowheads show where the blips of white noise occur. Subjects answer by placing a tick in the appropriate segment. In the earliest versions subjects were instructed to place a cross to indicate their choice; an observer suggested that ticks would be better, as there might be something inconsistent in asking children to indicate 'correct' items with a cross, which is always used as a symbol for 'wrong' by teachers. After the modifications described, the nature of the first answer blank used in schools was as below:-

8. 

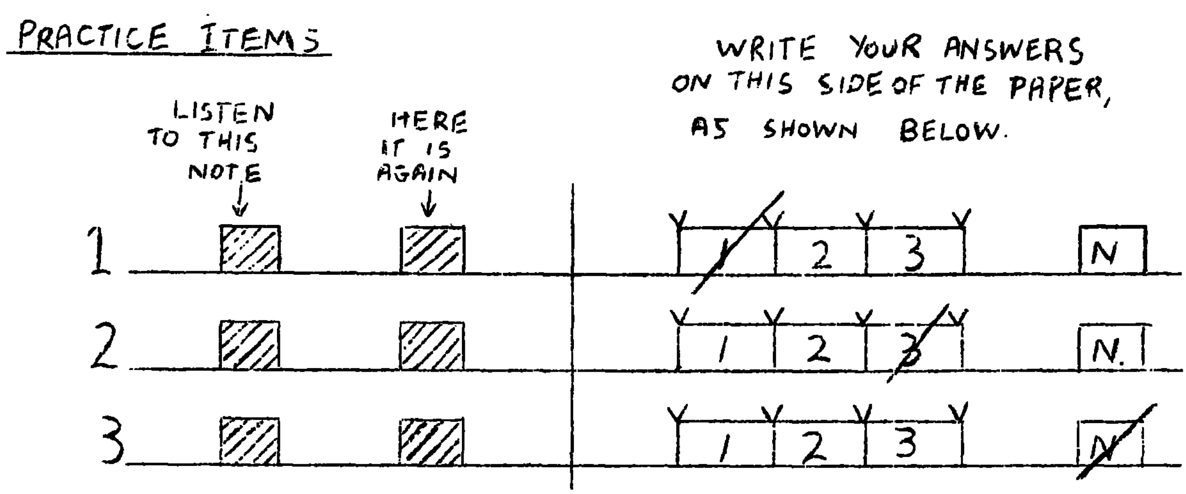
Here the number of choices in the glide tone is reduced to three, and a 'none' box has been added for stimulus tones not included in the glide tone. Answer blanks for other sub-tests were along similar lines.

Before administration to the Middlesbrough sample, three practice items were produced, for presentation before the test material itself. Overleaf is an illustration of the practice items, as they were represented on actual answer sheets. Note the minimal amount of verbal instructions.

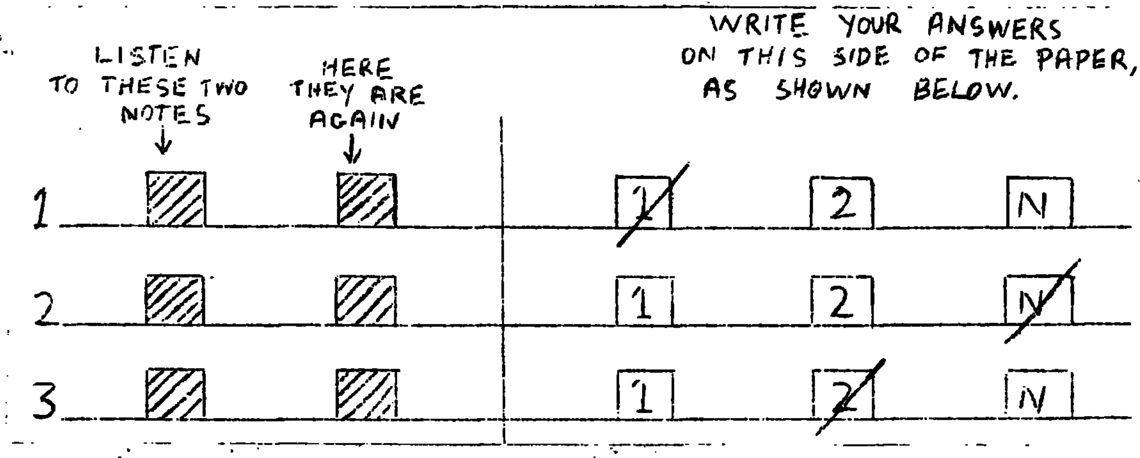
In this first pilot run, only the first five items from each sub-test were used, as the main aim was to discover if the tasks were meaningful to children.



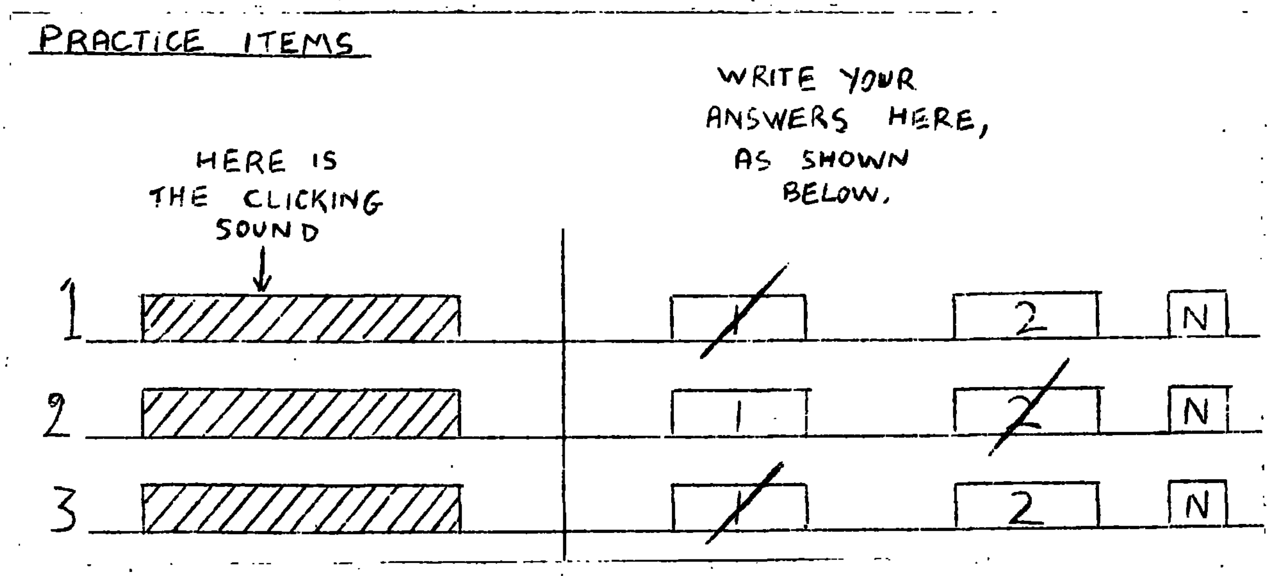
MELODY



PITCH.



AUDITORY TRANSPOSITION.(INTERVALS)



RHYTHM.

The Test Run.

In the form described above, the tests were administered to two classes of children in a primary school. Average age of both classes was 10 years, 3 months. There were 44 subjects. Instructions were given verbally; and at this stage questions were answered, which reduced the degree of standardisation of the instructions. In all later runs, instructions were always standardised.

Results.

There were no spoiled papers in this run. Scores and standard deviations for the parts and the whole were as below:-

Part 1:	Rhythm. Mean score, out of five:- 3.356	S.D. = 0.968
Part 2:	Auditory transposition. Mean score:- 3.556	S.D. = .856
Part 3:	Pitch. Mean score:- 4.022	S.D. = 0.857
Part 4:	Melodic Memory. Mean score:- 3.689	S.D. = .864
For total battery.	Mean score:- 14.804	S.D. = 2.195

Total possible overall was 20. Total possible on any sub-test was 5.

Conclusions.

The above results showed that a high mean score, of almost 75% was obtained. The low standard deviation showed that this short version of the test discriminated rather badly, but this was attributable to the small number of questions used, and copying on the part of subjects. The high average score, and the absence of any significant deficit on any one sub-test, suggested that the tasks were indeed meaningful, and that there was little difficulty in understanding the nature of the tests. On the basis of this test run, it was decided that a properly conducted run on a larger sample, and with completely standardised presentation would be worthwhile, using the unabridged version of the test. The fact that the results of this first full scale examination of the complete test's performance indicated that considerable modification was necessary merely illustrates how misleading the results obtained with the above group were. In fact, the results obtained on subsequent runs lead the writer to conclude that the classes used in this first pilot run were highly atypical. It seems possible that the Education Department, in their endeavour to secure 'good' classes for the experimenter, had selected classes of exceptionally bright children. Certainly the high mean scores of this group were not duplicated in the test runs immediately following. The conclusion that the task was meaningful was not in fact born out, and modifications to test material, and to administration, were made. Also, subsequent testing was performed in schools selected by the experimenter, in an endeavour to cover a more representative cross-section.

Section 3 shows the development of a pilot test battery, stemming from ideas presented in sections 1 and 2, up to the point where final evaluation of the testing method could begin. Section 4 is an account of the results obtained in administering the tests to over 2000 schoolchildren, and the changes that became necessary as a result; leading to the production of a test battery that met certain criteria.

SECTION 4.

Development of the Test Battery, leading to the Final Test Version.

First Test Run.

The version of the test described at the end of section 3 was administered to a sample of 117 children for the first test run. The subjects were from two different schools, situated in residential areas of Durham City. One school contributed classes of 8 - 9 (35 subjects) and 9 - 10 (27 subjects) years; classes of 9 - 10 (24 subjects) and 10 - 11 (31 subjects) came from the other school. There were 47 female and 70 male children in the sample. Standardised instructions were given on the tape; so the tests were self-administering in the same way as the Wing and the Bentley tests.

Reproduction was by way of a Tandberg tape recorder, model 12 - 21. This was situated centrally in front of the class. The nature of the tasks is such, however, that variations in the siting of the sound source, and variability in room acoustics, are extremely unlikely to influence the difficulty of the test material. (This is not true for other tests, however. The Seashore, and even the Bentley pitch test call for the perception of very small tonal differences, which might be influenced by interaction with external factors.)

Method.

Subjects were seated comfortably. The room was kept airy. Adjustments to seating were made wherever this was advisable, to eliminate cheating. Subjects were informed that they required only something to write with. Answer papers were distributed, usually with the help of 'volunteers'. The first task involved the filling in of a simple administrative form attached to the front of each set of answer papers. A sample of this form is shown below. All details except name, age and sex were written on the blackboard for the benefit of subjects who found difficulty in completing the form.

NEW TESTS OF MUSICAL APTITUDE

J. B. DAVIES. DEPT. OF PSYCHOLOGY, DURHAM UNIV.

PLEASE COMPLETE THE FOLLOWING:-

NAME _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

AGE yrs _ _ _ _ _ mnths _ _ _ _ _

TODAY'S DATE _ _ _ _ _ _ _

SEX _ _ _ _ _ _ _ _

SCHOOL _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

CLASS _ _ _ _ _ _ _ _ _ _

TOWN _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _ _

After completion of the administration form, subjects folded their papers over to the first test sheet. Each sub-test answer form was on a separate foolscap sheet. On this run, tests were in the order melodic memory, pitch, auditory transposition, and rhythm. Subjects were instructed to attend to the tape recorder, and told that all the necessary information was contained on the tape.

Before instructions for the first sub-test were given, there were some general instructions, to allow time for adjustment to the situation. These general instructions were as follows:-

> "Today, we are going to play a little game. I am going to ask you some simple questions about some noises that you will hear soon. Now, make yourself comfortable, pick up your pen, and be ready, because the game goes on fairly quickly. Ready?
>
> You answer by choosing one of the shapes on your answer paper, and crossing out the one you think is right. You will see how to do it from the practice items. Answer every question, even if you have to guess. No copying please. Your neighbours answer is probably wrong if it is different from your own! And remember, it's just a game, so do not worry".

Instructions for the specific sub-tests were as follows:-

Instructions for melodic (tonal) memory test.

"Here is part one of the test. In this part you have to remember some little tunes. First of all, you will hear a short tune. Some have three notes, like this. (Three note tune is played) Some have four notes like this. (Four note tune is played). You will hear each little tune twice, in case you did not hear properly the first time. After the second playing, you will hear a longer piece of tune, like this. (Six note tune is played) Say whether the longer tune has the little one somewhere in the middle of it. If the little tune is not in the longer one, answer 'N' for No. If it is, answer 'Y' for Yes. Now listen carefully. If you answer Yes, count how many notes are left after the little tune has stopped. There may be one, or two, or three, or four, or there may be no notes left over at all. Write how many notes are left over every time you answer Yes. Try the first practice item.

Here is the little tune. (three note tune) And again. (three note tune) Here is the longer tune. (six note tune, not containing shorter tune) The answer there was 'N' for No.

Here is the second practice item. Listen to the little tune. (four note tune) And again (four note tune). Now here is the longer tune. (eight note tune) The answer there was Yes; and one note was left over, so your answer should be 'Y' for Yes, and the figure 1.

Now try the last practice item on your own. (Complete item is played) The answer was 'yes', and one note left over. Now try the test, remembering that your answer will be 'N' for No, or 'Y' for Yes. Whenever you write 'Y' for Yes, your answer will be followed by the figure 1, or 2, or 3, or 4, or nought". (First sub-test commences.)

Instructions for Pitch Test.

"Here is part two of the test. In this part you have to say how high, or how low a note is. Listen to this note. (Stimulus tone is played). Here it is again, in case you did not hear it properly the first time. (Tone is played again.) Now listen to this strange sound. (Three segment glide tone is heard.) The note slides upwards; and you also heard four louder, shorter noises, or blips, dividing the sliding note into three sections. Follow them on your paper. The white shape shows the sliding note, and the arrowheads show where the blips come. Listen to the note at the beginning, and then say which part of the shape has that note in it. If you do not think any part has a note as high or as low as the one you heard, write 'N' for None.

Now try the first practice item, at slow speed. Practice item number one. Listen to this note. (Stimulus tone is played.) Here it is again (tone is repeated). Now......., is it here? (the first segment of the glide tone only is played). Or here? (the segment of the glide tone is played) Or here? (third segment is played) Or in none of them? The right answer was 1. Try it again but at the proper speed. Ready? (the complete item is played, but with the glide tone intact)

Now try the second practice item. (Complete practice item is played.) The answer was 3. Now try the third practice item. (Complete practice item is played.) The correct answer there was 'N' for None. Now try the test, remembering that your answer will be 1, or 2, or 3, or 'N' for None".

(Second sub-test commences)

Instructions for Auditory Transposition (interval) Test.

"Here is part three of the test. Listen to these two notes. (Stimulus interval is played.) The second note is a little bit higher than the first. Now listen to these two (a different pair of tones is played). The second note is a lot higher than the first. Listen to the notes, and remember how far apart they are. Then see if you can pick out which two notes that follow are the same distance apart. Like this.

Practice item number one. Here are two notes. (Stimulus interval is played.) And again, in case you missed them. (Stimulus tones are repeated.) The second note is just a little bit higher than the first. Remember how far apart they are. Now listen to these two notes. (First comparison interval is played.) And these two notes. (Second comparison interval is played.) Pick out the two notes that are the same distance apart as the first ones you heard. The right answer will be the two notes where the second one is only just a little bit higher than the first one. Listen to the first practice item again. Here are the two notes. (Stimulus interval is played.) And again, just in case you missed them the first time. (Stimulus interval is repeated.) They are very close together. Now..... are these two close together (first comparison interval is played? Yes. Are these two close enough? (Second comparison interval is played.) No. So you should cross out the square marked number one.

Try the second practice item. Listen to the two notes. (Stimulus interval is played.) And again. (Stimulus interval is repeated.) They are <u>very</u> far apart. Now, are these two notes far apart? (Second comparison interval is played.) No. Neither of them was far apart. So your answer would be 'N' for Neither.

Practice item number three. (The two presentations of the stimulus intervals are given.) Far apart. (The first comparison interval is given.) Close. (The second comparison interval is given.) Far apart. So number two is correct.

So what you have to do is listen to the first two notes, (you will hear them twice), and remember how far apart they are. Then you have to pick out the two notes that follow that are the same distance apart as the first ones. Now try the test, remembering that your answer will be one, or two, or 'N' for Neither". (Third sub-test commences)

Instructions for rhythm test.

"Here is part four of the test. Listen to this noise. (Six measures of triple time metre are played.) On every third click a bell rang. Listen again and count. (Metre is repeated.) Now listen to this. (Six measures of duple time are played.) The bell rang on every second click that time. Sometimes the bell will ring on every second click, sometimes on every third click, sometimes on every fourth, and sometimes on every fifth. Count, and decide which it is. After you have heard this sound, you will hear two short bursts of tapping. These are called rhythms. Say which rhythm fits with the clicking and ringing sound you heard at first, like this. Here is the first practice item. Here are the clicks. (Six measures of duple metre are heard.) The bell rang on every second click that time. Here is the first rhythm. (First comparison rhythm is heard.) And the second. (Second comparison rhythm is heard.) The first was the only one which would fit properly. Listen again, and count 'one... two... one... two' This may help you to see how the first rhythm fits. (Here four measures of duple time are played, with the sound of E's voice superimposed, counting 'one, two, one, two', etc.)

Try the second practice item. If the first rhythm fits, write the figure 1. If the second rhythm fits, write the figure 2. If neither fits, write 'N' for Neither. Practice item number two. Here are the clicks. (Six measures of quadruple metre are heard.) The bell rang on every fourth click then. The first rhythm. (First comparison rhythm is heard.). And the second. (Second comparison rhythm is heard.) The right answer was number two.

Now try the last practice item on your own. (Complete item is heard.) The right answer was 1. Now try the test, remembering that your answer will be 1, or 2, or 'N' for Neither.

(Fourth sub-test commences)

On all sub-tests, E merely calls out the number of each item before it commences.

Scoring.

On this version of the tests, scoring was as follows:-

Part One: Melodic (tonal) Memory. One point was scored for each correct choice of 'Y' (Yes) or 'N' (No). An additional point was scored for each correct insertion of a number, on items where 'Y' had been correctly selected. Since there were 7 correct 'Y' answers, total possible score was 22.

Part Two: Pitch. One point was scored for each correct selection of 1, 2, 3, or 'N' for None. Total possible 15.

Part Three: Auditory Transposition (intervals). One point was scored for each correct selection of 1, 2, or 'N' for Neither. Total possible 15.

Part Four: Rhythm. One point was scored for each correct selection of 1, 2, or 'N' for Neither. Total possible 15.

Total possible for whole test battery 67.

Results.

The data was examined primarily to discover the state of balance of the test. Mean scores for the parts of the test, and for the whole battery, are given below:-

Melodic (tonal) Memory.	Mean = 10.47.
Pitch.	Mean = 5.72.
Auditory Transposition.	Mean = 4.93.
Rhythm.	Mean = 4.945.
Total battery.	Mean = 26.06.

It is apparent that the obtained scores are rather low. This is indication that the tests as a whole are too difficult.

Below are given the difficulties of individual items in the sub-tests expressed as percentage of correct responses. N for each sub-test is also given. The difference between N for any sub-test, and the total sample size of 117 gives the number of spoiled papers.

Item Number.

SUBTEST	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M.M.	76	44.5	56	51.8	40.6	46.6	48.2	51.9	42.4	47.65	34	29.5	41.3	56.1	47
P.	13.8	54.9	22.25	38.1	42.2	49.6	62.2	15.93	24.4	20.1	39.1	35.93	32.8	39.1	81.2
A.T.	27.7	34	56.1	20.3	40.3	30.8	32.9	23.4	43.4	35	27.7	48.7	25.6	23.5	23.5
R.	51.2	5.9	15.4	16.4	38.6	33.3	37.5	45.9	22.8	27	40.7	22.8	29.1	41.7	45.9

M.M. = Melodic (tonal) Memory. (N = 110)
P. = Pitch. (N = 109)
A.T. = Auditory Transposition. (N = 105)
R. = Rhythm. (N = 111)

In constructing tests, a state of balance of the order of difficulty of the items is sought. Though there are no hard and fast rules, the following distribution of difficulty would normally be regarded as satisfactory in cases where the discrimination sought is based on the normal distribution:-

(a) About 20% of the items should range in difficulty from 0 - 40.

(b) About 60% of the items should range in difficulty from 40 - 60.

(c) About 20% of the items should range in difficulty from 60 - 90.

(d) Only a few very difficult and very easy items should be included. (Guilford, 1954.)

The scores obtained on the first test run are thus clearly too low.

Conclusions from the First Test Run.

Firstly, the pilot study on the Middlesborough children, with the short version of the tests, needs reconsideration, since the high scores obtained on that occasion were not reproduced on testing the larger sample. The conclusions that the test was meaningful, and that subjects readily understood the nature of the tasks also needs re-examination in view of the rather high number of spoiled answer papers. (10.25% on the auditory transposition test.) The results show that, though none of the sub-tests approaches the ideal very closely, the tests of auditory transposition and of rhythm are particularly unsatisfactory since the mean score on both these items failed to exceed the chance guessing score of five.

Balance of the tests is unsatisfactory; there are far too many items in the 0 - 40 range, and a deficit in the 40 - 60, and the 60 - 90 ranges. Modifications were therefore necessary; these took the following form:-

Modification to the first version of the tests.

In making these initial alterations to the test material, the basic layout, and mode of presentation, was preserved. The modifications at this stage consisted purely of changes to the material contained in individual test items. Redundant items in the 0 - 40 range were discarded. An attempt was made to replace them with easier items, by using existing items in the 40 - 60 and the 60 - 90 ranges of difficulty as a guide in the selection of new items.

Melodic Memory.

In this test, which was the most satisfactory of any, items 10, 14, and 15 were removed. These were replaced by one harder item, since the 0 - 40 range actually had a deficit of one item, and two easier items. It was hoped these adjustments would improve the balance. Also, one note was removed from all the comparison tunes, in an attempt to increase the average score by reducing the noise to signal ratio.

Pitch.

Here items 1, 4, 10, 11, 8, 9, and 13 were removed. An attempt was made to raise one very easy item into the 60 - 90 range, and the remainder of the changed items into the 40 - 60 difficulty range.

Auditory Transposition.

In this test, items 1, 2, 6, 7, 8, 13, 14, and 15 were removed. An attempt was made to produce three items in the 60 - 90 range, and five items in the 40 - 60 range. Changes made involved increases in the difference between 'correct' comparison tones and 'incorrect' comparison tones. The stimulus tones remained unchanged.

Rhythm.

Here items 2, 3, 4, 5, 6, 10 and 12 were replaced. Three items in the 60 - 90 range, and a further four items in the 40 - 60 range were required. No change was made to any of the metres, but the complexity of the comparison rhythms was reduced considerably in the items indicated.

No change was made in the instructions for any of the sub-tests. The changes made had no effect on the total duration of the battery, which remained at 36 minutes and 35 seconds, including instructions.

In its modified form, the test was given the title of the 'A' scale. In this form the tests were administered to a second sample of schoolchildren.

Performance of the 'A' scale:- Second Test Run.

Method.

The 'A' scale was administered to a sample of 691 subjects. These were schoolchildren ranging in age from 7 years to between 11 and 12 years. The subjects were from five schools in the Durham and Sunderland areas. Three schools were in the Durham city area, and one of these was closely annexed to a council housing estate. The remaining two were in outlying areas, in semi-rural communities. The sample comprised 47.76% males and 52.24% females. Conditions of administration, and instructions, were the same as for the first test run, within the limits possible.

Results.

No detailed results, such as sex differences, or performance of different age groups, are given for the second test run. It is apparent from examination of the overall means, and the item difficulties, that the adjustments made on the basis of the first test run, had failed substantially to effect the necessary changes in the performance of the tests.

Means and standard deviations for the parts of the test, and for the whole battery, are given below:-

Melodic (tonal) Memory.	Mean 11.863	S.D. 3.341
Pitch.	Mean 6.504	S.D. 2.894
Auditory Transposition (intervals).	Mean 4.747	S.D. 2.194
Rhythm.	Mean 6.162	S.D. 2.417
Total Battery.	Mean 29.211 (Variance 63.081)	S.D. 7.942

Comparison of the above with the results obtained on the first test run shows an increase in the scores on all parts of the test but auditory transposition. Differences between the means for melodic memory, pitch, and rhythm, on the two test runs were significant at the 1% level. Differences between the two runs for auditory transposition were not significant. Although the modified version shows some improvement in the mean scores, the magnitude of the improvement is too small; obtained means are still too low. Item difficulties for individual parts of the test are given below:-

Item Number.

SUBTEST	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M. M.	42.7	37.3	72.7	75.5	46.8	67.3	63.2	55.9	43.6	64.5	37.7	41.8	43.6	62.7	45.9
P.	20.2	61.5	26.6	48.6	46.8	51.4	52.3	14.7	32.1	22	48.6	50.5	40.4	50.5	84.4
A.T.	36.1	24.7	42.8	17.1	36.2	27.6	27.6	27.6	42.8	30.5	37.1	31.4	23.8	33.3	28.6
R.	78.4	28.8	30.6	36	37.8	54.9	57.6	49.5	37.8	20.7	32.4	20.7	28.8	53.2	48.6

For the total test battery, 12.7% of the papers were spoiled.

No further analysis of the data was carried out, since it was apparent that the tests were still not operating at a satisfactory level.

Conclusions.

From the low scores, and the imbalance in the item difficulties, it was apparent that the tests were still too difficult. In particular, the failure of the auditory transposition test to respond to the modifications made is particularly striking.

The number of spoiled papers suggests that too many subjects either (a) are unable to understand the nature of the tasks, or (b) due to the difficulty of the test items, become disconsolate, with lessening of attention and perhaps an increased tendency toward random response.

It should be born in mind that the sample used in the second test run comprised the entire age range 7 to 12 years. The first test run did not. Thus, the effects of the modifications made are likely to be attenuated to some extent by the lower scores of the younger age groups, and the possibly greater difficulty of understanding experienced by the younger children.

It was concluded that if the tests were to operate satisfactorily, considerable re-organisation of the test material was necessary; further minor modifications similar to the ones already carried out would be unlikely to bring about the desired improvements.

In addition to the above, certain behaviours exhibited by the children during administration of the test gave cause for concern. The most disconcerting was the grimacing, eye closing, and 'hands over ears' shown by a small percentage of subjects during the pitch test; discussion of the tests with subjects, and with class teachers, involving re-runs of certain parts of the tests, revealed that the source of discomfort lay in the white noise 'blips' used to break up the glide tones into their different segments. Clearly, some different signal was needed here.

Finally, the duration of thirty-six minutes was felt to be excessive especially for the very young children.

The main requirements for the new battery would be, therefore, as follows:-

1. The actual test items would need to be made easier on the whole.
2. Presentation of stimuli would need to be made clearer; to this end a general slowing up of the rate of presentation was thought to be desirable.
3. Instructions would need to be made clearer at certain points, and perhaps expanded at others.
4. In opposition to points 2 and 3 above, it was very desirable, for the younger children, that the duration of the test battery be cut to something in the order of twenty minutes.

The attempt to meet these somewhat conflicting demands, and the production of an extensively revised battery, titled form 'B', is now described.

The Revised Test Battery: Form 'B'.

In the construction of the new revision of the tests, no fundamental alteration to the nature of the tasks was made. Major changes, however, were made to the presentation of the test items, and narrower limits were set to the response categories.

Though the early experiences with the test material had not been particularly encouraging, there was no reason to suppose that the basic methods proposed could not be incorporated into a satisfactory test. A small experiment was carried out at this point which suggested that the basic methods might be along the right lines. This briefly consisted of the administration of the test form 'A' to a sample of 12 <u>adults</u>, five of whom were practising musicians. The mean total score for the musical group was 59 (range 54 - 65). For the non-musical group the mean score was 35.8 (range 29 - 47). Results for the sub-tests were as follows:-

	Subject No.	M.M.	Pitch.	A.T.	Rhythm.
Musicians.	1.	18	14	12	15
	2.	22	15	(5)	(12)
	3.	19	14	9	(12)
	4.	21	14	14	14
	5.	22	14	14	15
Non-Musicians.	6.	8	7	5	7
	7.	14	7	6	6
	8.	12	9	4	4
	9.	16	11	9	11
	10.	15	7	6	12
	11.	12	10	7	12
	12.	14	6	6	8

With the exception of the three results circled, the scores of the two groups on all parts of the test are mutually exclusive. The method then is capable of discriminating between criterion and non-criterion groups to a very high degree. Whilst the groups tested are exceedingly small, the results obtained added further evidence that the testing method could be made to work. Unfortunately, the need to produce a children's battery meant that no further, extensive, studies of the performance of 'A' scale with adult populations could be made. Two further minor studies using the 'A' scale will be referred to later. The 'B' form of the test, and the ways in which it differed from the 'A' scale will now be described.

Form 'B' of the Tests.

The four sub-tests used in the 'A' scale are retained, and keep the same names. Methods of producing 'statistical approximations to music', rhythmic and metre patterns, sine wave tones, and splicing of tapes, are all as for the 'A' scale.

Test of Melodic (tonal) Memory.

All the tones used in both stimulus and comparison tonal sequences were lengthened. In the 'A' scale, individual tones were of .53 seconds (4" at $7\frac{1}{2}$" per second). This was increased to .63 seconds ($4\frac{3}{4}$" at $7\frac{1}{2}$" per second). Inter stimulus time on the 'B' form was 2.26 seconds (17"), and time allowed for answering was 3.33 seconds (25"). To compensate for this increase in the length of items, it was necessary to remove the second (repeat) presentation of the stimulus tune; this resulted in an overall shortening of the duration of the test compared with the original 'A' form. This modification represents an attempt to 'trade' the advantages associated with slower rate of presentation (greater time allowed for consolidation) with the advantages associated with two presentations of the stimulus tune. (Strengthening of memory trace.)

The order of difficulty of the items was reduced by increasing the signal to noise ratio of each item considerably. This was done by shortening both stimulus and comparison tunes, as follows. Ten stimulus tones of three notes and five stimulus tones of four notes were used. This represents a considerable shortening of stimulus tones compared with the 'A' scale. The comparison tunes were also shortened; comparison tones of four and five notes were paired with the three note stimulus tunes. There were five of each. Five comparison tunes of eight notes were paired with the four note tunes. All the practice items comprised three note stimulus and five note comparison tunes. This shortening of the items, by reducing the number of tonal elements, is also advantageous from the point of view of the interference theory of forgetting. (Ceraso, J., 1967).

In addition, the material was so arranged that all items to which the correct answer was 'Y' for Yes would have either 1 or 2 notes left over. It was made absolutely explicit in the modified instructions, that each 'Y' answer would be always followed by either a figure 1 or a figure 2. This raised the guessing score to 9.125 out of 22.

Test of Pitch.

In this test, the procedure of reducing the number of presentations of the stimulus tones from two to one was again adopted. Though Wickelgren found differences in memory for the pitch of tones when these were increased in duration from 2 to 4 to 8 seconds, the differences, though consistent, were small. Also, the effects he observed were perhaps heightened by interference tones placed between stimulus and matching tones. The suggestion that short term memory for pitch decays more rapidly under certain types of interference, and the implication that greater exposure to the stimulus tones is necessary for a subject to perform accurately under greater degrees of interference, was born out by the following experimental study, performed by the author and J. Berriman. The study was carried out to find what differences in short term memory for pitch resulted when white noise, 'random' music, or silence, was interjected between stimulus and matching tones. The finding that recall is rendered worse by the interjection of material between stimulus and response is widely known.

(Conrad, 1959; Broadbent, 1958). There is also the implication that relevant interfering material is more damaging than irrelevant material. Below is given a very brief outline of the study performed:-

Method.

Two criterion groups were selected to perform the experiment:- a 'musically trained' group of 15 subjects who were members of the Durham University Music Department, and a 'musically untrained group' of 30 subjects with no musical training.

Apparatus.

The apparatus used was similar to that used in the pitch matching experiments. Both S and E were seated in front of audio-oscillators, producing sine-wave tones. S could be presented with a matching tone from E's oscillator; throwing a switch then enabled S to hear sound from his own oscillator, on which he had to match the stimulus tone. Both oscillators were balanced for output. Both S and E heard tones via headphones.

Between the stimulus and matching tones, E could interject a burst of white noise, or a sequence of random piano music, from a two-track tape recorder; or a period of silence.

Procedure.

S. was presented with a series of stimulus tones. Before being allowed to match these, he was subjected to a five second period of one of the three types of interference. Stimulus tone duration was 10 seconds. Controls, method of measuring responses, and procedure in the sound-proof room were the same as for the pitch matching experiments. Three stimulus tones were used. There were three conditions, and each tone was presented twice. Total number of trials per subject was therefore 18.

Results.

There was a large difference in the mean errors for the two criterion groups. As shown in the pitch matching experiments, the musical group proved far more accurate than the non-musical group. The effects of the three types of interference were as follows:-

Subjects.	Random Music.	White Noise.	Simple Delay.
Musical.	4.3	3.9	3.1
Non-Musical.	51.0	49.6	45.5

Average error, in cycles per second, for each condition.

Whilst an analysis of variance showed difference between groups significant at the 1% level, differences between conditions were not significant for either group. However, when differences between conditions are expressed as percentages of the error when only simple delay is present, the following differences emerge:-

1. Increase in error due to white noise delay; as percentage of error under simple delay;

   Musicians.............25.8%
   Non-Musicians......... 9.01%

2. Increase in error due to random music, expressed as percentage of error under simple delay;

   Musicians.............38.7%
   Non-Musicians.........12.08%

Differences between musicians and non-musicians are significant. It appears then, that though the base line performance of the musical group is superior to that of the non-musical group, the interjection of white noise or random music causes greater percentage performance decrement for the musicians than for the non-musicians. Under all conditions, the effects of the interference were in the same direction for both groups.

Conclusions.

The above is a very brief summary of the study. Only those results directly relevant to the present work are given. The implications of the study for the pitch test are as follows. Firstly, it is clear that the interjection of bursts of white noise in the comparison 'glide' tones does not merely serve as a cue to the start and finish of the different segments of the glide tones; it is almost certainly positively interfering for all subjects, and possibly more interfering for musical than non-musical subjects. Secondly, since longer presentations of stimulus tones are necessary when interference is present (Wickelgren), the need to keep the present test battery as short as possible suggests the need to keep interference between stimulus and response minimal.

In the revised version of the pitch test, only one presentation of the stimulus tone was given. Though this was not as long as the 8 second period necessary for complete consolidation of pitch found by Wickelgren, the minimal amount of interference between stimulus and response in the present test makes the use of shorter stimulus tones practicable. Also, it may well be that the subjects with the highest musical aptitude can internalise the pitch of the tone more rapidly than other subjects, and that this sub-test operates on this criterion. The glide tones were separated into their segments by simple silence, and the number of segments further reduced to two. The items were thus as follows:- after the announcement of the number of the particular item, the subject hears a tone of 3.2 seconds duration. After two seconds pause, the glide tone commences; after the first segment of the glide tone there is a silent period of 1.3 seconds, after which the second segment of glide tone is heard. 3.33 seconds are then allowed for answering. The subject answers 1, 2, or neither, depending on whether the stimulus tone occurs in the first segment, the second, or neither segment, of the glide tone. This reduced the number of alternative answers to three, raising the guessing score from 3.75 to 5.

Test of Auditory Transposition (intervals).

The length of the tonal elements was again increased, in exactly the same fashion as the melodic memory tests. The nature of the task performed by the subject was changed, in the following manner. Instead of hearing a stimulus interval, followed by a series of comparison intervals (as in the 'A' scale) the subject merely hears two consecutive intervals. There is no emphasis on selecting one of a series of comparisons. The subject listens to the two intervals, and is asked to say which pair of notes are the closest together, in a tonal sense. The exact nature of the task is made clear in the instructions.

On some items, both the intervals presented are identical musical intervals, in different keys. The subject is asked to write 1, or 2, to indicate which interval contains notes that are closer together; or to write 'same' if the two intervals are the same. This gives three alternatives, and a guessing score of 5. The actual difficulty of items was reduced by selecting intervals that had produced the greatest number of correct responses in the 'A' scale. Few difficult intervals, such as flatted fifths and flatted sixths, were included.

Test of Rhythm.

Mode of presentation was again changed. Instead of hearing two comparison rhythms following presentation of the metre, subjects now hear only one. They listen to the metre and the rhythmic pattern, and then merely answer 'yes' or 'no' to indicate whether or not the comparison rhythm has the same time signature as (i.e. 'fits' with) the metre. In order to reduce the guessing score from 50%, subjects were also asked to indicate whether the metre was duple, triple or quadruple, by writing 2, 3, or 4 on their answer sheets. One point was scored for each correct selection of one, two or three. Total possible score thus became 30, with a chance guessing score of 12.5.

An obvious point of contention in the above test, is that in asking subjects to indicate the nature of the 'metre' we are testing a counting ability, which has little to do with rhythmic ability; and therefore penalising those who have poor counting or numerical ability. Two studies were performed which suggest that this is not so.

Initially, from Piaget, 1941, we may note that a child aged 6 to 7 years usually has little difficulty in counting objects in excess of ten. The present task involves counting up to four only, but involves sonic rather than concrete objects. (for definition of 'sonic object' see A. Moles, 1968) The following study was performed, to test the hypothesis that the perception of the rhythmic metre was simply a numerical ability.

Method.

The sample comprised 127 children, all aged between 7 and 8. There was an approximately equal sex division. Subjects were asked to count various objects presented to them by the experimenter. A short answer blank was prepared, a sample of which appears below:-

1. HOW MANY BOOKS? ________ ☐
2. HOW MANY DOLLS? ________ ☐
3. HOW MANY TAPS? ________ ☐
4. HOW MANY WHISTLES? ________ ☐
5. HOW MANY SLAMS? ________ ☐
6. HOW MANY SNIPS? ________ ☐
7. HOW MANY JARS? ________ ☐
8. HOW MANY BOXES? ________ ☐

The objects used in the study presented various types of information, all demanding the abstraction of number. The highest number used was five.

Procedure.

All the objects were placed in a large cardboard box, so they were not visible to subjects until produced. Before each item, E made the following speech:-

"I am going to show you some things, which I want you to count. Now! How many (books, dolls)?"

The objects used, and the types of information they provided, were as follows:-

1. Books (3). These were held high in the air. Visual objects.
2. Dolls (3). These were held high in the air. Visual objects.
3. Taps (4). E taps firmly on the desk with a ruler. Subjects close their eyes. Auditory only.
4. Whistles (2). E blows firmly on an Acme whistle. Subjects close their eyes. Auditory only.
5. Slams. (1). E holds up a book, and slams it closed vigorously.. Visual and auditory object, abstract.
6. Snips (5). E holds up a pair of kitchen scissors, and snips them vigorously. Visual and auditory object, abstract.

7. Jars (2). E holds up, and shakes, some small jars containing a few dried peas, which rattle. Visual and auditory, concrete.

8. Boxes (4). E holds up, and shakes, some small boxes containing a few dried peas which rattle. Visual and auditory, concrete.

Results.

Out of the whole sample, only one incorrect result was obtained. This was an error on item 8, which was scored 3 instead of 4.

Conclusion.

The counting of various types of objects up to the number of 5 was performed with extreme ease by a sample of seven year old children. Effectively, there were no errors. The much higher number of errors obtained on the 'counting' part of the rhythm tests suggests that more than simple counting is required, and that differences between scores on the rhythm tests are unlikely to be due to differences in counting ability up to the number 5.

The second study described here is an investigation into short term memory and rhythmic abilities, supervised by the writer, and carried out by V. Pomfret, 1969. Only one aspect of the study will be referred to here; namely, an investigation to show if the factor of 'general rhythmic ability' postulated by Thackray was simply a general short term memory factor.

Method.

Four groups of sixteen subjects were selected randomly from the age groups 5, 6, 7 and 8 years. Each subject was given four tests. Two of these were tasks in short term memory, involving digits and words respectively. The other two were 'rhythm' tasks involving "clapping the rhythm of a melody", as described by Thackray, and "marching on the table with both hands, in time to the 'music".

Results.

The results obtained in the study were extremely detailed; only the broad findings and conclusions are given here. The main findings were, from an examination of partial correlations, that clapping the rhythm of a tune did not depend only on short term memory, but involved other things as well. Rhythmic performance of the 'marching in time to the music' type did not correlate at all with short term memory. It was also found that performance at the short-term memory tasks correlated significantly with intelligence. but that performance on the rhythmic items did not correlate with intelligence.

Conclusions.

The main point from these two studies, as far as the present work is concerned, is that ability to perform well on the rhythm tests proposed is unlikely to be influenced to any significant degree by counting abilities, or by general short term memory, or by intelligence.

Thehypothesis that the 'metre' part of the test introduced into the 'B' form is simply a test of counting, or of short term memory, is not substantiated, therefore. In the revised form, all quintuple items were removed, to reduce the number of more difficult items. The complexity of certain of the comparison rhythmic patterns was also reduced, using easier items from the 'A' scale as a guide. The form of the rhythm tests was now as follows. The subject hears six measures of a metre, in duple, triple, or quadruple time. This is followed by a pause of 3.2 seconds (24") during which S writes the figure 1, 2, or 3. The comparison rhythm then follows, after which subjects must say whether it was in the same time as the metre, answering 'Y' for Yes or 'N' for No. Time allowed for answering was 3.33 seconds.

Modifications to the Test Answer Blank.

Examination of the spoiled papers from the 'A' scale results showed that some subjects found the test blank confusing. There was a tendency for answers to be placed inside shaded blocks, other than the answer blocks. The additional information provided by the pictorial representations of all the events in any item was clearly misleading to some subjects. On the 'B' scale, all blocks representing stimulus events were therefore deleted. The blocks pertaining to comparison or 'response' events were retained. Subjects thus answered by selecting one of the shapes; there were no 'non-response' shapes. A sample of the 'B' form of the answer sheet is given overleaf. No further modifications were made to the answer blank, which now proved satisfactory.

Modifications to the Instructions.

Though remaining basically the same in form, the verbal instructions were modified slightly to fit in with the new test situations. Certain 'doubtful' points were clarified. The general instructions were the same as for the 'A' scale. Instructions for the specific sub-tests were as follows:-

Instructions for Melodic (tonal) Memory.

"Here is part one of the test. In this part you have to remember some little tunes. First of all you will hear a short tune. Some have three notes, like this. (Three note tune is heard) And some have four notes like this. (Four note tune is heard)

After the little tune has stopped, you will hear a longer piece of tune like this. (Six note tune is heard) Say whether the longer tune has the short one somewhere in the middle of it. If you _do_ hear the short tune somewhere in the middle of the longer one, answer 'Y' for Yes. If you do _not_, answer 'N' for No. Now listen carefully. If you have answered 'Y' for Yes, count how many notes were played after the little tune had stopped. There will be either one note, or two notes, left over every time the answer is Yes, so write how many notes are left over every time you choose 'Y' for Yes. At the top of the page there are three practice items with the answers already filled in for you. After you have heard them carry on answering in the way shown. Here is the first practice item. Here is the little tune. (Three note tune is heard) Here is the longer one. (Four note tune is heard) The little tune _was_ in the longer one, and one note was left over at the end; so your answer would be to cross out the letter 'Y' for Yes, and write the figure one. This has already been done for you in the first practice item. Here is the second practice item. Listen to the little tune. (Three note tune is heard) Here is the longer one. (Five note tune is heard) The little tune was not in the longer one, so the answer would be 'N' for No. Here is the last practice item. (Complete item, with three and five note tunes, is heard) The answer there was Yes, and two notes were left over. Now try the test, remembering that your answer will be 'N' for No, or 'Y' for Yes. Every time you choose 'Y' for Yes, your answer will be followed by the figure 1, or the figure 2. (Test commences)

Instructions for Pitch Test.

Here is part two of the test. In this part you have to say how high or how low a note is. Listen to this note. (Tone is heard) Now listen to these two strange sounds. (Two-segment glide tone is heard) You heard two notes, both sliding upwards. This is what you have to do. Listen to the note at the beginning, and remember it. Then listen to the two sliding notes. Say which of these has the first note in it. If you think the first sliding note is correct, answer by crossing out the number one. If you think the second sliding note is correct cross out the number two. If you do not think that either of the two sliding notes have a sound as high or as low as the one heard at first, choose 'N' for Neither. Here is the first practice item. Here is the note. (Tone is heard) Now, is it here? (First segment of glide tone is heard) Or here? (Second segment of glide tone is heard) Or in neither of them? The right answer was number one. Try the second practice item. Here is the note. (Tone is heard) Is it here? (First segment of glide tone is heard) Or here? (Second segment of glide tone is heard) Or in neither of them? It was in neither of them, so your answer would be 'N' for Neither. Now try the third practice item. (Complete item is heard) The right answer there was number two. Now try the test, remembering that your answer will be one, or two, or 'N' for Neither. (Test commences)

Instructions for Auditory Transposition (intervals) Test.

Here is part three of the test. In this part you have to say how far apart two notes are. Listen to these two notes. (A melodic interval is heard) The second note was just a little bit higher than the first. Now listen to these two notes. (Interval is heard) They were much further apart. The second note was a lot higher than the first. Here is what you have to do. First of all you will hear two notes. Remember how far apart they are. Then you will hear two more notes. Remember how far apart they are. You have to pick out the two notes that are the closest together. If the first two notes are closer together than the second two, answer one. If the second two notes are the closest, answer two. If both pairs of notes are the same distance apart, answer 'S' for same. Here is the first practice item. Here are the first two notes. (Interval is heard) They were very far apart. Here are the second two notes. (Interval is heard) Much closer together. So the answer is two; the second two notes were much closer together than the first ones. Here is the second practice item. (Interval is heard) Close. (Interval is heard)

Far apart. So the right answer would be one. Here is the third practice item. (Interval is heard) Close. (Interval is heard) Close. They were both close together; so the answer would be 'S' for same. Now try the test, remembering that your answer will be one, or two, or 'S' for Same. (Test commences)

Instructions for Rhythm Test.

Here is part four of the test. Listen to this noise. (Six measures of triple time metre are heard). You heard a clicking sound, and a bell rang on every third click. Listen again, and count. (A further six measures are heard, with E's voice superimposed, counting one, two, three, etc.) Now listen to this. (Six measures of quadruple are heard) The bell rang on every fourth click that time. Sometimes the bell will ring on every second click, sometimes on every third click, and sometimes on every fourth click. Count, and decide which it is. After the clicking and ringing sound, you will hear a short burst of tapping. Say whether this tapping sound fits with the clicking and ringing sound you heard first. This is what you have to do. Listen to the clicks and the bell, and decide whether the bell rings on every second click, or every third, or every fourth. Decide which it is, and write two, or three, or four in the first square. Then listen to the tapping sound, and say whether it would fit with the clicking sound, answering 'Y' for Yes, or 'N' for No. Here is the first practice item. Here are the clicks. (Six measures of metre are heard) The bell rang on every second click, so the figure two should go in the first square. Can you see the figure two on your answer sheet? Now, here is the tapping sound. (Rhythmic pattern is heard) Did it fit with the clicking and ringing sound? The answer was 'Y' for Yes. Here is the second practice item. Here are the clicks. (Six measures of metre are heard) The bell rang on every fourth click. Here is the tapping sound. (Rhythmic pattern is heard) The answer there was 'N' for No. Here is the third practice item. (Complete item is heard) The answer there was the figure three, and the letter 'Y' for Yes. Now try the test, remembering that your answer will be the figure two, or the figure three, or the figure four, followed by 'Y' for Yes or 'N' for No. (Test commences)

Administration.

This was the same as for the 'A' scale, with the following addition. As a visual aid, a large version of the practice items was constructed for each sub-test. These were drawn on cards size 1' 8" by 2' 6". E displayed these cards when the appropriate instructions were being reproduced. A selection of ticks and numbers was made, cut out of card. These symbols were affixed to the larger sheets as the instructions directed. This additional aid seemed particularly helpful to the younger children. Duration of the total test battery was now 30 minutes 15 seconds.

The first version of the 'B' form was administered to a further sample of schoolchildren, in the form described.

The 'B' form. First Test Run.

The tests were administered to 448 schoolchildren, in the age range 7 to 11 - 12 years. The sample comprised 46.8% males and 53.2% females. Four schools from the Durham City, Leamside, and Sherburn areas provided the sample. One of the schools (Durham Bluecoat School) was a C. of E. school and did not necessarily draw its children from the immediate neighbourhood; though the fact that it was a day attendance school restricted its hinterland effectively to the Durham City area. (No residential schools were used in any of the test runs) Appearances suggested that the schools covered a wide socio-economic cross-section. With the exception of the visual-aid described above, administration of the first version of the 'B' form was the same as on previous runs, within the limits possible.

Results.

Mean scores and standard deviations for the sub-tests and the whole battery are given below:-

<u>Melodic (tonal) Memory</u>.	Mean 11.86	S.D. 2.823
<u>Pitch</u>.	Mean 6.9	S.D. 2.68
<u>Auditory Transposition</u>.	Mean 6.43	S.D. 2.154
<u>Rhythm</u>.	Mean 18.19	S.D. 5.8
<u>Total Battery</u>.	Mean 43.4	S.D. 8.208

(Expected guessing score for whole battery 31.6)

Item difficulties for the different sub-tests are given below:-

Test Items.

SUB TEST	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M.M.	75.7	41.6	71.5	65.5	51.8	69.5	68.9	37.3	39.9	50.1	35.6	58.7	52.7	42.4	47.6
P.	43.7	56.5	24	50.5	48	59.9	58.2	34.3	38.5	52.2	60.3	41.1	42.8	38.5	40.3
A.T.	53	58.1	40.9	64	324	58.9	53.8	504	22.1	42.7	39.2	22.1	34.2	23.9	46.9
R.	61.5	57.2	47.8	58.1	61.5	59.8	65.8	52.1	70	66.6	55.5	62.4	64.1	64.9	64.1

<u>Spoiled Papers</u>.

Down from 12.7% to <u>3.125%</u>.

<u>Conclusions</u>.

Differences between means for the sub-tests and the whole battery on the 'A' and 'B' scales are all highly significant,* with the exception of the melodic memory test. The most significant effect of

* When comparing the means, scores were reduced to a percentage, on the rhythm test, and the total battery, in order to make the scores comparable.

the 'B' form however is the drastic reduction in the number of spoiled papers. This suggests that the new form of the test was more readily understandable, and that random response due to lapse of attention was reduced.

The balance of item difficulties on the 'B' form was very different from balance on the 'A' form. The preponderance of difficult items in the 0 - 40 range has disappeared. Below is given a table of the numbers of items occurring in each of the difficulty categories; results for the second version of the 'A' scale are given in brackets. Figures for the rhythm and auditory transposition tests are particularly interesting.

% Difficulty.

SUBTEST		0 - 40		40 - 60		60 - 90	
		'A'	'B'	'A'	'B'	'A'	'B'
	M.M.	(2)	3	(7)	7	(6)	5
	P.	(5)	4	(8)	10	(2)	1
	A.T.	(13)	6	(2)	8	(0)	1
	R.	(9)	0	(5)	6	(1)	9

Ideally, there should be about 3 items between 0 - 40, about 9 items between 40 - 60, and about 3 between 60 - 90. Examination of the above table shows that the balance of tests one and two approximates roughly to the desired distribution. Test three, which was grossly imbalanced in the 'A' scale, is improved, but still has far too much ceiling. The rhythm test is transformed, however, from a test with too much ceiling into one with far too little ceiling; the balance on the 'B' form is almost a 'mirror image' of the balance on the 'A' scale.

Comparison of the two test forms reveals a slight reduction in the standard deviations for sub-tests in the 'B' form. Though this is difficult to interpret, the cause is almost certainly the small increases in the possible guessing scores. In view of these increases to the guessing scores, means and standard deviations are remarkably stable. This perhaps suggests that little systematic guessing was going on, but that wrong answers were genuine false positives. It may also indicate that the difficulty of the various tasks is not unduly influenced by the mode of presentation.

Before detailed item analysis, some minor adjustments and a further test run were thought necessary.

The 'B' Form. Final Test Run.

Adjustments were made to items in the four sub-tests on the basis of the data obtained from the last test run. No further changes were made to instructions, test blanks, or to the mode of presentation of the material. A complete description of the tests in their final form is given in appendix one. A description of the 'A' scale is also given, which serves for comparison.

The final version of the 'B' form differed from the earlier version in two ways. Firstly, the scoring on the rhythm test was altered. Instead of one point being scored for each correct selection of 2, 3, or 4, plus an additional point for each correct selection of 'yes' or 'no', one point was awarded for each correct 'yes' or 'no' and this was made contingent upon correct selection of a number. The total possible score thus now became 15. This change in the marking served two purposes. On the first run of the 'B' form, mean score overall was 60.6%. The guessing score was 47.7%. By altering the scoring system, the overall mean score is reduced, and the difference between obtained mean and 'guessing' mean is increased. Thus with the new scoring method, mean obtained score was 41.6%, and the guessing mean was 16.7%. Secondly, the new system of scoring is perhaps logically more apt. Under the old system, a person who wrongly perceived the initial metre still had a 50% chance of scoring one point on the 'yes' 'no' part of the test even though a correct answer to the second part presupposes correct perception of the metre. Under the new system, correct perception of metre is made a condition for scoring a point on the second part of each test item.

The second change involved the transposition of the entire test into a lower 'key'. (All items were moved down the frequency spectrum by the same amount percent) Bentley (Musical Ability in

Children) and Cleall, C. (The Natural Pitch of the Human Voice) have both shown the effects of change of range on pitch judgements. In the present case the findings of Bentley are confirmed, in that scores are improved as a result of transposition of items into the natural range of the voice; the findings of Cleall were used as a model for transposition in the present case.

Finally, changes were made to individual items in the various sub-tests in an attempt to improve the balance of the difficulty of items.

Method.

The test was administered to a sample of schoolchildren in the age range 7 to 11 - 12 years. Details of the sample are given below:-

Total number in sample: ....537.

Numbers in each age group:

7 - 8 years	106.	(N.B. children were grouped according to actual age, and not merely by average age of class or grade.)
8 - 9 years	145.	
9 - 10 years	139	
10 - 11+ years	147	

Sex distribution: ....46.18% male.

....................53.82% female.

The schools providing the sample were all in the Durham City and surrounding areas. Four schools were involved.

Administration, and the use of a visual aid, were exactly as on the previous test run, within the limits possible.

Results.

Means and standard deviations for the individual sub-tests and for the whole battery are given below:-

1. Melodic (tonal) Memory ....Mean 11.73.....S.D. 3.071
2. Pitch......................Mean 7.78.....S.D. 2.682
3. Auditory Transposition......Mean 7.31.....S.D. 2.784
4. Rhythm......................Mean 7.16.....S.D. 3.267

Total Battery...............Mean 33.93.....S.D. 8.516

(0.05 discrepancy due to rounding errors)

'Guessing score' for the whole battery........21.6

Note that for this final version of the tests, possible total scores for the sub-tests are:-

1. Melodic (tonal) Memory...........22
2. Pitch............................15
3. Auditory Transposition...........15
4. Rhythm...........................15

Total............................67

Item difficulties for all items in the different sub-tests are given below:-

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
M.M.	84	40	78	63	40	72	64	29	48	53	24	53	55	49	46
P.	64	50	44	67	49	63	52	37	52	46	42	37	68	50	42
A.T.	61	55	59	62	28	63	51	55	47	41	58	52	31	26	43
R.	49	31	41	42	50	44	61	39	59	61	36	57	48	53	42

Examination of the balance for the sub-tests shows that all are reasonable approximations to the desired distribution. Below is given a table of the distribution of item difficulties. Ideal distribution is given in brackets; plus any relevant comments.

	0 - 40		40 - 60		60 - 90		Comments.
	Obtained	Ideal	Obtained	Ideal	Obtained	Ideal	
M.M.	3	(3)	7	(9)	5	(3)	Two of the items in the 60 - 90 range have ratings of 63 and 64.
P.	2	(3)	9	(9)	4	(3)	Two items in the 40 - 60 range are rated 42. One item in the 60 - 90 rated 64.
A.T.	3	(3)	9	(9)	3	(3)	
R.	3	(3)	10	(9)	2	(3)	One item in the 40 - 60 range has a rating of 59.

It will be noted from the above that in those cases where the 3-9-3-distribution did not obtain, a change of only two or three percent on two items at the most would be sufficient to bring about the ideal balance of difficulty. The obtained balance is thus fairly satisfactory.

In order to determine the effects of age on test scores, mean scores for the sub-tests and for the whole battery were computed for each age group. Note that the age groups used are 'real' age groups and not class or grade groups. Thus, the effects of brighter children who may be a year ahead, or less bright children who are retarded a year, do not confound the results. Means for the different age groups are given below, for sub-tests and for the total battery:-

	M.M.	P.	A.T.	R.	Total Battery
7 - 8 years.	9.765	6.41	6.65	5.6	29.253
8 - 9 years.	10.96	7.35	7.21	6.72	32.24
9 - 10 years.	11.506	7.805	7.33	7.49	34.131
10 - 11+ years.	12.77	8.68	7.99	8.33	37.77

The results show small but consistent increases in scores on all parts of the battery, as a result of increasing age.

In terms of mean scores and balance, the test may be regarded as being fairly satisfactory. All means for all age groups are above guessing scores, but there is still ample 'ceiling' in the test to permit high scorers to be picked out.

Conclusions.

Though far from being perfect, the means for the different age groups and the distribution of item difficulties suggest that in its present form the test battery operates in a fairly consistent and meaningful way. Further test runs on very large samples would certainly improve the battery, and produce more data in terms of which its performance can be reviewed. The final test run produced the following 'standard' scores. For each age group, and for each condition, three figures are given. The central figure is the mean, and the two flanking figures give the scores which cut the distribution of the subjects into equal thirds. (Tertile scores) The figure in the red square is the only one which is in any way anomalous. (It ought to fall between 6.22 and 6.86)

	M.M.	P.	A.T.	R.
7 - 8 years.	8.78 9.765 11.18	5.69 6.41 7.88	6.08 6.65 8.21	4.16 5.6 7.67
8 - 9 years.	9.92 10.96 12.67	6.67 7.35 8.57	6.22 7.21 8.52	6.07 6.72 8.73
9 - 10 years	10.43 11.506 13.25	6.91 7.805 9.33	5.9 7.33 8.9	6.6 7.49 9.7
10 - 11+ years.	11.6 12.77 14.36	7.58 8.68 10.22	6.86 7.99 10.28	7.34 8.33 10.28

Standard Scores:- Means and Tertiles.

It is possible now to interpret scores on the tests in a meaningful way by comparing obtained scores with the standard scores.

The major part of the experimental work, involving the selection and organisation of certain types of material, and bringing these to a point at which further large scale sampling would be valuable, have been described. It remains now to describe the results of an individual item analysis, and introductory studies of reliability and validity, the results of which show how further development of the test material might take place, and the direction which future research might take.

PERFORMANCE OF THE TEST BATTERY.

Form 'B'. Item Analysis of the Final Test Run.

The following simple item analysis was performed on the results obtained from the above test run. The method involved the calculation of the double tetrachoric coefficient of correlation between each individual item and total scores on the sub-test involved. The method used is outlined in P. E. Vernon, (1948). The tertile scores for each sub-test are calculated. Then for each item, two correlations are calculated. One compares scores in the top two tertiles, or in the lower tertile, with right or wrong for each particular item. The other compares scores in the bottom two tertiles, or the top tertile, with right or wrong for the particular item. This produces a 'chi-squared' type of layout for each coefficient. For each item, two values of rt are obtained. Comparison of the twovalues obtained shows whether the item operates better in the upper or the lower part of the scale. The average of the two obtained coefficients gives an overall index of discrimination for the individual item. Number of calculations performed was 2 x 4 x 4 x 15 = 480, separate sets of coefficients being obtained for all age groups. It should be noted that the cosine-pi tables for the rapid determination of the tetrachoric correlation coefficient (M. D. Davidoff and H. W. Goheen, 1953) are accurate only when variables are dichotomosed at the

ITEM ANALYSIS.

PART ONE.

MELODY.

ITEM No.	AGE 7-8				AGE 8-9				AGE 9-10				AGE 10-11				MEANS			
	U	L	Dr	dif	U	L	Dr	dif	U	L	Dr	dif	U	L	Dr	dif	Mean U	Mean L	Mean dif	Mean Dr
1	.22	.54	.38	-.32	.5	.5	.5	0	.17	.55	.36	.38	.74	.75	.745	-.01	.4	.585	-.18	.496
2	.33	.02	.175	.31	.40	.28	.34	.12	.62	.35	.475	.29	.21	.42	.31	-.21	.39	.26	.13	.325
3	.69	.78	.735	-.09	.76	.61	.685	.15	.65	.83	.74	-.18	.78	.55	.66	.23	.72	.69	.03	.705
4	.63	.55	.59	.08	.58	.73	.655	-.15	.35	.47	.41	-.12	.58	.55	.56	.03	.535	.575	-.04	.55
5	.47	.46	.465	.01	.13	.20	.165	-.07	.26	.25	.255	.01	.46	.74	.6	-.28	.33	.44	-.08	.371
6	.01	.41	.21	.40	.26	.51	.385	-.25	19	.32	.255	-.13	.71	.51	.61	.2	.29	.44	-.145	.365
7	.30	.49	.395	-.19	.37	.30	.335	.07	.68	.64	.66	.04	.77	73	.72	.04	.53	.54	-.01	.535
8	.37	.28	.325	.09	.5	.39	.445	.11	48	.40	.44	.08	.57	.36	.465	.21	.48	.36	.12	.42
9	.28	.14	.21	.14	.43	.47	.45	-.04	.44	.42	.43	.02	.64	.54	.59	.1	.45	.39	.06	.42
10	.06	.23	.145	-.17	.54	.32	.43	.22	.56	.57	.567	-.01	.65	.41	.53	.24	.45	.38	.07	.415
11	.5	.13	.315	.37	0	0	0	0	.12	.14	.13	-.02	.25	.35	.3	-.1	.22	.155	.063	.187
12	.35	.12	.235	.23	.22	.38	.30	-.16	.25	0	.125	.25	.30	.06	.18	.24	.28	.14	.14	.21
13	.22	.16	.19	.06	.50	.61	.55	-.11	.12	.17	.145	-.06	48	.05	.26	43	.33	.25	.08	.29
14	.31	.19	.25	.12	.43	.30	.365	.13	.66	.62	.64	.04	.62	.74	.68	-.12	.505	.46	.043	.482
15	.48	.43	.455	.05	.36	.24	.30	.12	.54	.51	.525	.03	.58	.78	.68	-.2	.49	.49	0	.49

U = dichotomisation at upper tertile.

L = dichotomisation at lower tertile.

Dr = double tetrachoric correlation coefficient.

dif = difference between U and L.

PART TWO.

PITCH.

	AGE 7-8				AGE 8-9				AGE 9-10				AGE 10-11				MEANS			
ITEM No.	U	L	D rt	dif	U	L	D rt	dif	U	L	D rt	dif	U	L	D rt	dif	Mean U	Mean L	Mean dif	Mean Drt
1	.54	.37	.455	.17	.47	.47	.47	0	.56	.56	.56	0	.41	.45	.43	-.04	.495	.462	.032	.478
2	.26	.34	.3	-.08	.22	.26	.24	-.04	.71	.6	.655	.11	.57	.54	.555	.03	.44	.435	.005	.437
3	.58	.45	.515	.13	.59	.57	.58	.02	.54	.56	.55	-.02	.48	.32	.4	.16	.547	.475	.072	.511
4	.25	.60	.425	-.35	.06	.36	.21	-.3	.49	.44	.465	-.05	.52	.41	.465	.11	.33	.452	.123	.391
5	.39	.11	.25	.28	.62	.47	.545	.15	.56	.47	.515	.09	.58	.65	.615	-.07	.537	.425	.113	.481
6	.23	.39	.31	-.16	.34	0	.17	.34	.39	.35	.37	.04	.87	.62	.745	.25	.457	.34	.118	.398
7	.58	.57	.575	.01	.63	.43	.53	.20	.77	.65	.71	.12	.67	.59	.63	.08	.663	.56	.103	.611
8	.4	.24	.32	.16	.56	.49	.525	.07	.21	.08	.145	.13	.43	.55	.49	-.12	.4	.34	.06	.37
9	.27	.2	.235	.07	.27	0	.135	.27	.37	.19	.28	.18	.61	.41	.51	.2	.38	.2	.18	.29
10	.18	.52	.35	-.34	.41	.42	.415	-.01	.25	.25	.25	0	.28	.23	.255	.05	.28	.355	.075	.318
11	.53	.41	.47	.12	.48	.47	.475	.01	.54	.57	.555	-.03	.46	.49	.475	-.03	.503	.485	.018	.494
12	.5	.52	.51	-.02	.24	.60	.42	-.36	.26	.32	.29	-.06	.58	.56	.57	.02	.395	.5	.105	.448
13	.42	.73	.575	-.31	.47	.49	.48	-.02	.6	.49	.645	.11	.54	.22	.38	.32	.508	.483	.025	.495
14	.37	.27	.32	.1	.34	.56	.45	-.22	.58	.53	.555	.05	.63	.74	.685	-.11	.48	.525	.045	.503
15	.67	.4	.535	.27	.42	.50	.46	-.08	.61	.66	.635	-.05	.64	.25	.445	.39	.585	.453	.133	.519

PART THREE.

AUDITORY TRANSPOSITION. (INTERVALS)

Item No.	AGE 7-8 U	L	D rt	dif	AGE 8-9 U	L	D rt	dif	AGE 9-10 U	L	D rt	dif	AGE 10-11 U	L	D rt	dif	MEANS Mean U	Mean L	Mean dif	Mean D rt
1	.44	.45	.445	-.01	.78	.84	.81	-.06	.85	.75	.80	.1	.81	.87	.84	-.06	.72	.728	.008	.724
2	.17	.31	.24	-.14	.52	.40	.46	.12	.56	.61	.585	-.05	.80	.74	.77	.06	.512	.515	.003	.514
3	.54	.74	.64	-.2	.64	.62	.63	.02	.62	.57	.595	.05	.71	.67	.69	.04	.628	.65	.023	.638
4	.21	.34	.275	-.13	.47	.30	.385	.17	.39	.3	.345	.09	.86	.66	.755	.21	.483	.398	.085	.44
5	.30	.50	.40	-.2	.35	.41	.38	-.06	.64	.71	.675	-.07	.5	.36	.43	.14	.448	.495	.048	.471
6	.36	.36	.36	0	.23	0	.115	.23	.20	.23	.215	-.03	.25	.25	.25	0	.26	.21	.05	.235
7	.56	.34	.45	.22	.76	.49	.625	.27	.63	.76	.695	-.13	.92	.67	.795	.25	.718	.565	.153	.641
8	.72	.32	.52	.40	.85	.80	.825	.05	.75	.62	.685	.13	.71	.48	.595	.23	.758	.555	.203	.656
9	.63	.57	.6	.06	.49	.50	.495	-.01	.57	.58	.575	-.01	.65	.7	.675	-.05	.585	.588	.003	.586
10	.55	.39	.47	.16	.53	.51	.52	.02	.62	.50	.56	.12	.18	.51	.345	-.33	.47	.478	.008	.474
11	.16	.27	.215	-.11	.53	.35	.44	.18	.49	.33	.41	.16	.63	.31	.47	.32	.453	.315	.138	.384
12	.68	.69	.685	-.01	.51	.50	.505	.01	.64	.73	.685	-.09	.71	.8	.755	-.09	.635	.68	.046	.658
13	.36	.32	.34	.04	.16	.43	.295	-.27	.46	.19	.325	.27	.46	.32	.39	.14	.36	.315	.045	.338
14	.47	.33	.40	.14	.29	.35	.32	-.06	.36	.16	.26	.20	.78	.73	.755	.05	.475	.393	.083	.434
15	.17	.24	.205	-.07	0	0	0	0	0	0	0	0	.2	.4	.3	-.2	.093	.16	.068	.126

PART FOUR.

RHYTHM.

	AGE 7-8				AGE 8-9				AGE 9-10				AGE 10-11				MEANS			
ITEM No	U	L	D rt	dif	U	L	D rt	dif	U	L	D rt	dif	U	L	D rt	dif	Mean U	Mean L	Mean dif.	Mean D rt
1	.63	.79	.71	-.16	.52	.40	.46	.12	.42	.49	.455	-.07	.41	.57	.49	-.16	.495	.563	.068	.529
2	.7	.61	.655	.09	.58	.58	.58	0	.56	.46	.51	.1	.55	.33	.44	.22	.597	.495	.103	.546
3	.59	.72	.655	-.13	.51	.42	.465	.09	.77	.81	.79	-.04	.53	.61	.57	-.08	.6	.64	.04	.62
4	.55	.34	.445	.21	.47	.59	.53	-.12	.61	.78	.695	-.17	.41	.3	.355	.11	.51	.503	.008	.506
5	.75	.69	.72	.06	.62	.61	.615	.01	.64	.76	.7	-.12	.7	.71	.705	-.01	.677	.693	.015	.685
6	.27	.76	.515	-.49	.50	.35	.425	.15	.48	.63	.555	-.15	.64	.57	.605	.07	.473	.578	.105	.525
7	.51	.62	.565	-.11	.36	.72	.54	-.36	.6	.67	.635	-.07	.38	.41	.395	-.03	.463	.605	.143	.538
8	.3	.64	.47	-.34	.34	.36	.35	-.02	.44	.37	.405	.07	.72	.7	.71	.02	.45	.518	.068	.484
9	.59	.64	.615	-.05	.44	.37	.405	.07	.43	.35	.39	.08	.64	.68	.66	-.04	.52	.51	.01	.517
10	.75	.73	.74	.02	.72	.89	.805	-.17	.84	.76	.80	.08	.54	.66	.6	-.12	.713	.76	.048	.738
11	.60	.68	.64	-.08	.41	.47	.44	-.06	.6	.47	.535	.13	.24	.38	.31	-.14	.463	.5	.038	.481
12	.72	.71	.715	.01	.62	.7	.66	-.08	.46	.58	.52	-.12	.7	.47	.585	.23	.625	.615	.01	.62
13	.52	.47	.495	.05	.57	.43	.5	.14	.48	.47	.475	.01	.84	.67	.755	.17	.603	.51	.093	.556
14	.53	.69	.61	-.13	.41	.49	.45	-.08	.61	.66	.635	-.05	.5	.76	.63	-.26	.513	.65	.13	.581
15	.49	.67	.58	-.18	.34	.56	.45	-.22	.55	.5	.525	.05	.44	.18	.31	.26	.465	.47	.023	.466

median. The error is only slight for near-median splits but becomes appreciable when one or both variables are dichotomised far from their medians. Error also increases with the size of r cos-pi for any given condition of unequal dichotomisation. Corrections to all the coefficients obtained were made with the help of correction graphs. (Norman C. Perry, Norman W. Kettner, Alfred F. Hertzka, and Eugene A. Bouvier, 1953) All coefficients given have been corrected for non-median dichotomisation. A copy of the correction graphs is given overleaf. The graphs given are for use when p1 or p2, but not both, is greater than 0.5. Different graphs are used where these conditions do not obtain. Results from the item analysis are given on the pages following the correction graph.

Item Analysis: Discussion of Results and Conclusions.

Though double tetrachorics, and means, are given in the preceding tables for all items, a meaningful interpretation can only be made by examining the upper and lower tetrachoric coefficients. No conclusions on the performance of individual items can be made from examining differences or double rt's alone, since results can cancel each other out. Though there are slight differences between the numbers in different age groups, all coefficients of 0.25 or over may safely be regarded as significant at the 1% level or higher

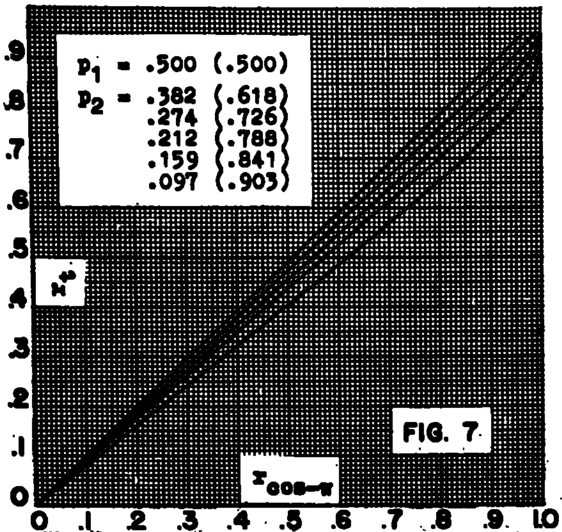


FIG. 7

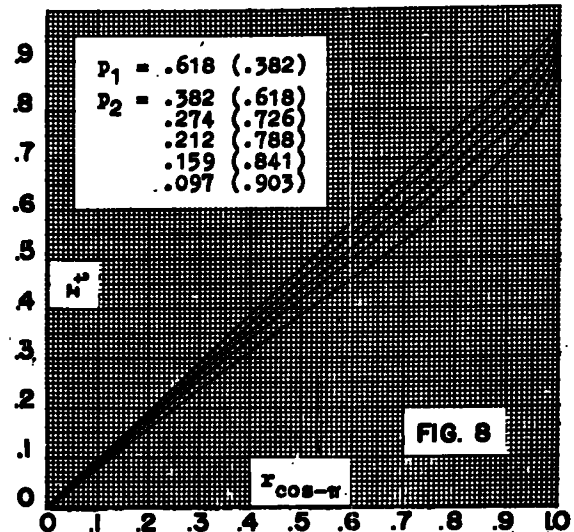


FIG. 8

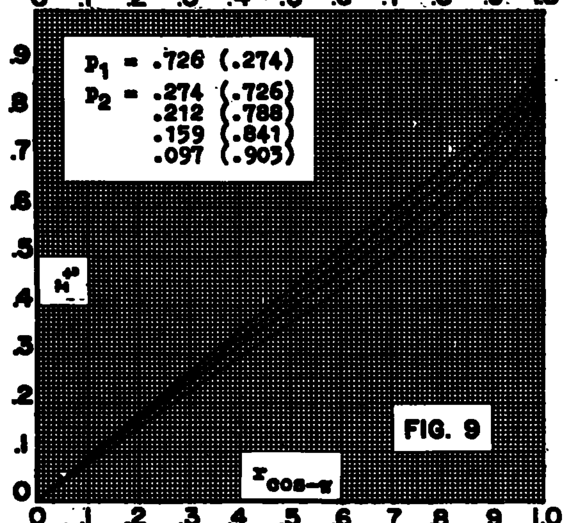


FIG. 9

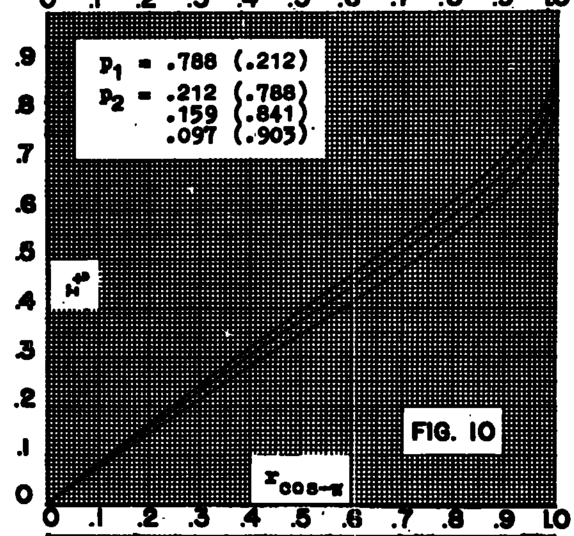


FIG. 10

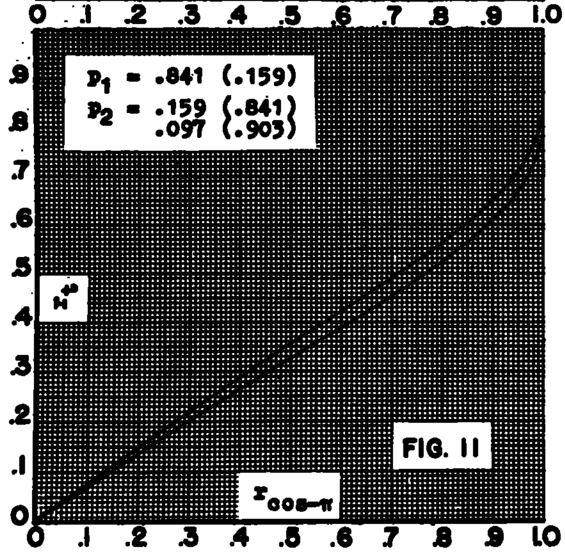


FIG. 11

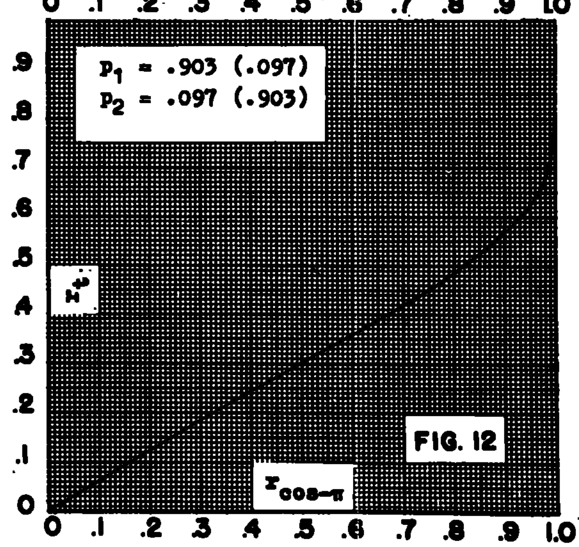


FIG. 12

The data shows that the majority of items have good or fair discrimination in both the upper and lower ranges; on the whole, discrimination seems to be slightly better in the upper range than the lower. Note that on test batteries of this type, low scores are always likely to be less reliable than high scores. With the exception of the items indicated on the results sheets, all items have a discrimination of at least 0.3 in one range. The majority exceed this value in both ranges. Two items show very poor discrimination. These items, in the red squares, are item 11 on the melodic (tonal) memory test, and item 15 on the auditory transposition test. The latter item is particularly bad; but its late position in the test might exaggerate the weakness. In addition, two further items, in the green squares, are picked out as being weak, in that coefficients do not exceed 0.3 on either upper or lower parts of the scale. It should be noted that removal of either of the very poor items would have an adverse effect on the balance of both sub-tests. Also, it is not necessarily beneficial to have several items with extremely high discrimination, as this leads to redundancy in the remaining items. (General Ref. Thorndike, R. L., 1949)

The main conclusion from the item analysis is that performance of most items justifies their inclusion. There is room for considerable doubt in the case of four items. Two of these are particularly bad, and should be replaced before further widespread testing. On the whole, the results are remarkably consistent; examination of the mean U and L scores for the different age groups show small but consistent increases in the power of discrimination with increasing age.

Form 'B'. Reliability of the Test Battery.

Two assessments of the reliability of the battery were made. The first involved calculation of correlation coefficients from scores obtained from a sample of 97 children tested on two separate occasions. The second involved administration of the battery to 118 children, with subsequent analysis using the Spearman-Brown formula for split half reliability. There was a certain discrepancy in the results obtained with these two analyses; but it should be remembered that the Spearman-Brown is more a measure of adequacy of item sampling, and that temporal stability does not enter into this statistic. The temporal stability of the test was measured by administering the tests to classes of school-children on two separate occasions. Over one hundred were tested on each occasion, but absentees reduced the number who performed the test twice to 97. Four months and one week elapsed between the two trials. The group tested included subjects from all age groups. The coefficients obtained for the sub-tests for the whole age range are not as high as one would like them. This is due to two reasons. Firstly, examination of the data for individual age groups shows a steady decrease in the size of the obtained coefficients of reliability with decreasing age; this probably reflects the greater difficulty of the younger children in maintaining attention, or the greater liability to fluctuations of motivational state. Secondly, the use of a normal sample, rather than a high motivational criterion group, lowers the coefficients. Reliability figures would be considerably improved if higher age groups only were used; (Bentley's reliability data was obtained from a sample with a mean age of 10 years 9 months. Range was from 9 years 10 months to 11 years 9 months.) and the use of high motivation groups would certainly improve the figures.

Results obtained, are given below, in two tables. The first table shows results obtained with a sample of 37 children in the top two age ranges.

Reliability of Sub-tests, and whole battery, for age ranges 9 - 10 years, and 10 - 11+ years.

Part one. Melodic (tonal) Memory.	0.76
Part two. Pitch.	0.58
Part three. Auditory Transposition (intervals)	0.54
Part four. Rhythm.	0.63
Reliability of total battery	0.82

Below are the reliabilities for the younger age groups. (N = 60)

Reliability of Sub-tests, and whole battery, for age ranges 7 - 8 and 8 - 9 years.

Part one. Melodic (tonal) Memory.	0.73
Part two. Pitch.	0.412
Part three. Auditory Transposition (intervals)	0.37
Part four. Rhythm.	0.46
Total Battery.	0.701

The reliability for the older children, taking the battery overall, is fair; and approaches the figures found for the Bentley and the Wing tests. Note that the upper age group in the present study has a lower mean age than in Bentley's study. Figures for the younger children, especially on certain sub-tests, are rather poor. In the absence of data on the reliability of the Bentley and other tests when used with very young children, it is difficult to make any comparative statements; it is very likely that other test batteries would suffer in a similar way with these age groups.

With both groups, the reliability of the melodic (tonal) memory test is quite good; and with the younger children actually exceeds the reliability of the whole battery. This may be due to the more involved scoring system on this part, which gives a total out of 22, rather than 15. In a way this may be equivalent to lengthening the test; increasing length has a beneficial effect on reliability coefficients. Also there may be a primacy effect.

In conclusion, the data on the reliability of the testsis rather limited. However, there is little doubt that reliability falls off quite markedly when the test is used with younger children. With children in the age range 9 years to 11+ the present battery compares favourably with the Bentley, the reliability data for which was obtained with slightly older children.

Note that product-moment coefficients were used for all the above; some workers have used rank-coefficients, which take into account only relative positions of testees' scores. No account is taken of magnitude of difference between scores; results from the two methods are discrepant. Product-moment coefficients make fullest use of the data available.

As an indication of the adequacy of item sampling, split half coefficients were calculated for all sub-tests separately. All age groups were analysed together. The coefficients obtained from the Spearman-Brown formula are given below:-

Split-half Reliability of the Sub-tests.

Part one. Melodic (tonal) Memory.	0.64
Part two. Pitch.	0.86
Part three. Auditory Transposition (intervals)	0.64
Part four. Rhythm.	0.91

Item sampling is therefore fair for parts one and three, and good for parts two and four.

Conclusions from Reliability Studies.

For older children, aged 9 years to 11+ years, the reliability figures obtained for the present battery are comparable with those found for the Wing and the Bentley tests. Reliability of the sub-tests was rather poor for younger children, however. It would appear essential, therefore, that all workers should state clearly what sample was used in the derivation of their reliability coefficients, as these are probably subject to effects similar to the ones found in the present study. It also seems clear that the criterion for accepting or rejecting certain levels of reliability should vary in stringency for different age groups; it is certainly easier to obtain high values with older populations, and an acceptability level of about 0.8 for adults may well be a less stringent criterion than a reliability coefficient of 0.7 obtained with young children. Clearly, it is desirable however that workers should attempt to maximise reliability under all conditions.

The observed effects of age are probably due to the greater fluctuation in attentional and motivational states in the younger children, and no test of the present type is equipped to compensate for these things.

Results from the split half reliability study indicate that item sampling varies from adequate to good on different parts of the test.

Form 'B'. Validity of the Test Battery.

The validity data on the present test battery is rather sparse. A thorough examination of the validity of the measures is a separate study in itself. However, a certain amount of data on test performance has been obtained, results from which are discussed below. First, however, certain difficulties associated with the assessment of validity are discussed, with special reference to the present study. Generally speaking, three broad methods are available for validating test material of the present type. These are: 1. validation longitudinally, that is over a considerable time, against a criterion closely tied to the construct the test is intended to measure. 2. comparison of the scores obtained by so-called 'criterion' and 'non-criterion' groups. 3. examination of the correlation between test performance and the ratings of 'experts' or people presumed to be able to 'rate' in a meaningful way. The following problems arose from consideration of these methods. In the present study, concerned primarily with the selection and development of test material, method 1. (longitudinal validation) is clearly impossible within the scope of a three year study.

Alternative two (comparison of scores for criterion and non-criterion groups) is not entirely satisfactory. In brief, the validities obtained from such studies tend to be in excess of those obtained when the test is given to normal non-selected samples. This is partly due to the increased motivation of the criterion as opposed to the non-criterion groups. There are other important reasons why cross-validation between criterion and normal samples may be unsatisfactory. (For detailed account see Meehl and Rosen) Briefly, the value of a test is very closely related to the nature of the sample to be tested. Unfortunately, in the present study, no figures for the penetration of musicality (according to some criterion) in the population as a whole are available. Nevertheless, the following argument might apply. For instance, the penetration of schizophrenia in the normal population is about 0.85%. If a clinician always diagnoses 'not schizophrenic' he will be correct 99.15% of the time. Suppose, by comparing criterion and non-criterion groups, he finds that a certain test picks out 80% of the schizophrenic group as being schizophrenic. On the face of it, the test is a most useful tool. However, if he diagnoses from this point entirely in accordance with the test, he will make 20% errors. Previously, by guessing 'non-schizophrenic,' his error was only 0.15%. Now, he incorrectly classifies as schizophrenic 20% of the non-schizophrenic population. Thus, in the above hypothetical case, the test will only be a useful tool where comparison between schizophrenic and non-schizophrenic groups is necessary. With an unselected sample, the clinician will be more accurate if he merely guesses 'non-schizophrenic'. Similarly, with tests of musical ability, any validity data obtained from comparison of criterion and non-criterion groups can only be interpreted usefully <u>if the base rates for certain levels of musical ability are known</u>. Any attempt to determine these levels would be most useful. In the absence of such data, validity results from this type of cross-validation do not give any clues as to the overall usefulness of the test.

Alternative three, validation against teachers' rankings, also presents certain difficulties. The basic assumption is that teachers' rankings are in some way a useful criterion. Yet what do we conclude in the case where a child who is rated as being very mediocre or poor by his teachers scores very highly on an

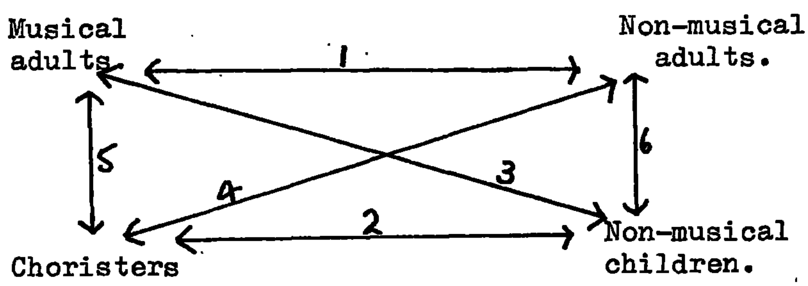
I.Q. test? Cronbach and Meehl argue as follows (1955) "Teacher judgements once constituted the criterion against which the individual intelligence test was validated. (Note the use of the past tense:- author) But if today a child's I.Q. is 135 and three of his teachers complain about how stupid he is, we do not conclude that the test has failed. Quite to the contrary, if no error in test procedure can be argued, we treat the test score as a valid statement about an important quality, and define our task as that of finding out what other variables-personality, study skills, etc. modity achievement or distort teacher judgement". In the present case, the value of this method of validation is further called into doubt by the overall lack of agreement between the rankings of various teachers. Correlations between rankings obtained from various teachers were almost without exception insignificant, and coefficients themselves were low. The question arises as to the value of validity coefficients obtained with the use of some criterion which is itself unreliable.

The above paragraphs underline briefly the major problems confronting the present investigation. None of the problems has in fact been solved. Data, obtained from comparison of various small criterion and non-criterion groups, and from the rankings of one teacher, are given below. The only defence which it is possible to offer is that the objections raised above apply equally to the results of other workers. Results are also given from a comparison of the Bentley Tests and the tests of the present writer; but since the Bentley tests are themselves validated in large part against the type of criteria described, it is difficult to interpret the results obtained.

Overleaf are the results obtained from four small samples. The groups consist of 1.20 choristers from the Durham Choir School (average age 11 years 4 months), 2.8 musical adults with performing skill as criterion. (These were all unconnected with the University) 3. 11 non-musical adults (unconnected with the University) 4. 13 randomly chosen children with average age 11 years and one month. For each group, means and standard deviations for the sub-tests and for the total battery are given.

It will be observed from these figures that both musical groups are superior to both non-musical groups on all parts of the test. Also, musical adults are slightly superior to musical children, but non-musical children score slightly higher than non-musical adults. The theory that musical abilities may decay through lack of use after children leave school is well known, though the present finding cannot really be regarded as confirmation. A short series of t-tests was performed on the total scores of the various groups. The results of the comparisons, and a key showing which comparisons were performed, are given below:-

Key to t-tests.



Results of t-tests.

Comparison 1. (M.A./N-M.A.) t = 6.934 significant at better than .001

Comparison 2. (C./N-M.C.) t = 4.615 significant at better than .001

Comparison 3. (M.A./N-M.C.) t = 6.506 significant at better than .001

Non-Musical.	Adults. N = 11				
M.M.	P.	A.T.	R.	Total.	
13.0	8.909	6.909	8.571	38.091	MEANS
2.558	2.152	1.929	2.922	6.28	S.D.

Musical Adults.	N = 8				
M.M.	P.	A.T.	R.	Total.	
20.4	13.75	11.5	13.625	59.46	MEANS
1.9	1.479	3.122	1.576	6.265	S.D.

Musical Children.	Choristers. N = 20				
M.M.	P.	A.T.	R.	Total.	
15.65	11.75	12.25	10.75	51.45	MEANS
3.9	2.98	2.277	2.624	7.606	S.D.

Non-Musical.	Children. N = 13				
M.M.	P.	A.T.	R.	Total	
12.9	8.71	8.4	9.01	39.02	MEANS
2.559	2.559	2.801	2.5	6.877	S.D.

Comparison 4. (C./N-M.A.) $t = 4.805$ significant at better than .001.

Comparison 5. (M.A./C.) $t = 2.546$ not significant (p greater than .01)

Comparison 6. (N-M.A./N-M.C.) $t = 0.328$ not significant (p greater than 0.6)

All differences between musical and non-musical groups are therefore highly significant. Differences within the non-musical groups are not, though the difference between musical adults and choristers is significant at better than the 5% level.

Finally, a comparison of two samples of college students was performed by D. Hargreaves using an earlier version of the tests. He found that a sample of students studying music scored significantly more highly than a sample of students not studying music. There were 12 musical and 28 non-musical students involved in the study. Differences between the mean scores (47.3 for the musical group; 33.67 for the non-musical) were highly significant.

The attempt to validate the tests against the rankings of teachers was almost entirely unsuccessful. This was largely due to the lack of agreement shown by the teachers who supplied rankings. Even simplified rankings, such as picking the 'best' five children and the 'worst' five, failed to produce any consistent results. The ways in which music is taught in many junior schools in the area is probably responsible for this. Two situations are frequently encountered. Firstly, there is often no specialist music teacher in the schools, and several teachers may take the music sessions. Their perception of a child's performance is invariably clouded by general behavioural and performance criteria not related to musical performance. Secondly, the teaching methods are often not suited to the observation of individual differences in performance; frequently, music lessons consist entirely of listening to classical records, (so-called 'appreciation' sessions. The problems of 'appreciation' in the absence of any basic foundation in terms of wider musical activities have been discussed earlier.) or group singing. In this latter category, the endless repetition ('rehearsal') of hymn tunes for morning assembly seems as limiting in terms of 'musical exploration' as it is widespread. There were notable exceptions, where Orff type instruments were in use, in conjunction with certain excellent B.B.C. broadcasts; and composition and individual or small group performance encouraged. Such instances were very rare, however. (Only two out of all the schools tested could be described as progressive in the approach to musical education.) It is possible that the North East is particularly poorly endowed in the above respects, or that the sample chosen was atypical. However, certain other areas do seem to be rather better organised as far as the musical education of the young is concerned.

A rank coefficient was calculated which showed close agreement between test performance and general musical performance, for the choristers of the Durham Choir School. The cathedral organist and choir master supplied a tentative ranking. The close association between Conrad Eden and the musical lives of the choristers suggests that this ranking would be more reliable than most of the other rankings attempted. Spearman's rank correlation method yielded a coefficient of 0.68, between rankings and test performance. Dr. Eden showed great enthusiasm and subsequently expressed a desire to use the test as an aid in selection of future choristers.

Lastly, a comparison was made between scores obtained on the Bentley tests and the present battery. Both tests were administered to some 30 children in the 9 - 10 age range. Absentees reduced the number who worked both tests to 25 subjects. Product moment coefficients were calculated for the totals of both test batteries, and also for the different sub-tests. Results are given below:-

Comparison of Bentley and Davies testing systems.

Correlations Obtained.

1. Davies v Bentley 'melodic memory' (tonal) memory)......0.297
2. Davies v Bentley 'pitch'...............................0.42
3. Davies v Bentley 'rhythm'..............................0.256
4. Davies 'auditory transposition' v Bentley 'chords'.....0.39
5. Davies total battery v Bentley total battery...........0.657

None of the coefficients for the sub-tests is significant except that for the pitch test, which is just significant at the 5% level. The coefficient for the total batteries is highly significant, at better than the 0.1% level however.

Test form 'B'. Conclusions from the Validity Data.

Three types of study have been attempted. For teacher's rankings the data is very sparse indeed. The point has been made, that in the ultimate analysis, only a longitudinal study can provide really definitive data about test validity. The problems associated with other procedures have been briefly outlined. Data from other sources is presented above; notwithstanding the difficulties described, sufficient evidence exists to suggest that the test can be described with caution as 'valid'. In particular, the results from cross-validation using groups, show complete separation of musical from non-musical subjects, independent of age. Since in the large non-selected samples used in the main test runs, very high scores occurred in small but consistent numbers, independent of musical training, it seems unlikely that the test discriminates simply in terms of musical tuition. The results obtained from the one ranking study do not merit further comment. The comparison with the Bentley measures shows that, with both tests in toto, results are in fairly good agreement; sufficient to suggest that both batteries are measuring the same thing overall, at least to some extent. The low correlations obtained between the four sub-tests, however, suggest that these are on the whole not comparable. This is not surprising when the differences in the nature of the test material are considered. It is interesting to speculate where the overall agreement of the total batteries comes from, in the absence of any close agreement in the subtests.

<u>Footnote</u>:

General references for the methods employed in determining reliability, validity and norms for the final version of the test battery included Phillips, B. N. and Weathers, G. (1958), Lorge, I, and Thorndike, R. L. (1957), Cureton, E. E. (1950), Cronbach, L. J. and Meehl, P. E. (1967), Ebel, R. L. (1961) and others. Since these all deal with fairly general issues and procedures, no specific references are given in the text.

SECTION 5.

Conclusions from the Study.

Sections 1 and 2 of the present work represent a general coverage of work in the field of musical ability, leading to the selection for further study of certain types of material. Sections 3 and 4 are concerned solely with the experimental development of the selected material, leading to the production of a test battery. A better perspective of the work overall is obtained if certain of the goals stated or implied in sections 1 and 2 are re-stated in the light of the experimental findings.

Initially, it must be stated that the value of the work, if any, lies not simply in the production of a different test battery. A fairly comprehensive array of test material is already available, and the addition of one more battery to those already existing is not in itself particularly meaningful. The results are likely to be of more use as sources of information about musical perception, testing techniques, and the application at a very rudimentary level of signal detection and information theory in a field where such applications have been somewhat sparse. By applying simple, purely psychological, criteria to the selection and development of certain material, it has been shown that in the field of musical measurement such methods are not inappropriate; within certain limits, the test works, and compares favourably with other batteries. The point is that the approach adopted is shown to be a reasonable one in the light of the results obtained. Secondly, if the type of test material chosen does operate in a satisfactory way, we are in a position to say more about the nature of basic musical abilities; to see what is fundamental to certain tasks, and what merely 'noise in the system'. The Mursell type of view, that musical aptitude can only be measured in a thoroughgoing musical situation needs, at least, clarification.

Section 1 of this work dealt in some detail with factorial studies of musical ability, using data obtained from certain better known test batteries. The point was also made regarding the drawbacks associated with variables which are factorially over-complex. Guilford has underlined the need for greater simplicity; "rotations and interpretations would be much simplified if each variable were of complexity one; that is, if it measured only one common factor to any appreciable extent". Though the simplicity of the present data would probably not normally merit such detailed analysis, a simple factor analysis was performed on the results obtained from the present tests. (Thurstone's centroid method) Before giving the results of this simple analysis, the reader is referred to page 39 of this thesis, and the following passage. "It would be advantageous, therefore, if by manipulating short-term memory variables we could produce a test battery which, in view of Guilford's comment, produced substantially only one common factor".

A matrix of the correlations between the sub-tests is given below:-

	M.M.	P.	A.T.	R.
M.M.		0.356	0.269	0.256
P.	0.356		0.362	0.3125
A.T.	0.269	0.362		0.283
R.	0.256	0.3125	0.283	

Inter-sub-test Correlations.

Analysis of the above produced the following main factor loadings for the four sub-tests.

Melodic (tonal) Memory.............0.524
Pitch..............................0.645
Auditory Transposition.............0.551
Rhythm.............................0.50

No further analysis was carried out, asthe above accounted for rather more than 95% of the common variance.

From the above, it is clear that inter-sub-test correlations are on the whole rather small. Insofar as the sub-tests do measure common variables, the extraction of the one main factor accounts for virtually all the variance. The factor loadings themselves are moderate rather than high. The conclusion from the above must be that on the whole the sub-tests measure fairly specific aspects of musical ability; under normal circumstances the size of the obtained inter-correlations would hardly merit a factorial study. It is interesting that insofar as the tests do measure common or group factors, one single factor accounts for most of the variance. In view of the way in which the test has been constructed, it is tentatively suggested that such a common factor is concerned with short term memory for musical and quasi-musical materials.

A detailed study of the relationships between the present test battery and other abilities would be a considerable help in the task of more positive identification of the common factor in the test battery. At present such a body of information is not available. There are a few helpful indications from certain of the studies so far described, however. A brief comparison of classes from schools in which 'streaming' effectively takes place shows that there are small but not significant differences between streams. Very high or very low scores seem equally likely to occur in either group. The highest score obtained was from an eight-year old girl from the lower stream in a school at Leamside. This school was notable for the very low social and economic status of many of the children, and the rather low rate of examination success achieved. Teachers frequently found results surprising, especially in cases where 'bright' children achieved low scores.

Also, a study has been described showing the absence of any significant relationship between intelligence and scores in simple rhythm tests. The factor analysis shows comparable loadings on the common factor, for all parts of the test, implying that the other sub-tests do not measure intelligence to any appreciable extent. Finally, if the sub-tests were substantially tests of intelligence, one might have expected higher inter-sub-test correlations.

An interesting study was performed by D. Hargreaves to show the relationship between scores on the 'A' scale of the tests, and the type of 'creative intelligence' (divergent production abilities) discussed by Getzels and Jackson (1962), Wallach and Kogan (1965), MacKinnon (1962) and others. Hargreaves devised various scales for measuring divergent musical thinking, involving re-arrangement of elements, number and uniqueness of responses, detection of similarity, auditory imagery, and devising questions to test various aspects of appreciation. The musical material used was varied, ranging from 'pop' to modern jazz and classical. No frivolous material was included. His findings to some extent parallelled those of Getzels and Jackson, but in a specifically musical context. Performance on the 'divergent' tests he devised showed little relationship with scores on the 'convergent' test of musical aptitude. We can fairly confidently conclude that the present test battery does not measure creative talent in any way. In the musical sphere, as well as in the field of more general 'intelligence', creativity must be treated as something separate from the simple convergent skills involved in tests of the present type, and different ways for measuring, and encouraging, its development must be evolved. (Terman, L. M., 1954, Mednick, S. A., 1962)

A point emerging from examination of data on other test batteries is the vital need to specify at all times the age group from which data was obtained. This is of the utmost importance for producers of children's tests, but less importance for adult tests. Where tests are designed for administration to children whose abilities and natural talents are still developing, different results are obtained for different age groups. In the present study, it has been noted how mean scores increase throughout the age range. This is easily anticipated, and test constructors invariably give age norms rather than overall norms. However, it is necessary to go beyond this. In the present work, we have observed how the reliability of the tests, which is comparable with other batteries in the upper age range, falls to a very disappointing level for the younger children. In a similar way, validity is also a function of the age range used; fluctuations in attention and motivation being primarily responsible. In this respect, the use of older children, or even adults, and criterion/non-criterion groups in the derivation of validity data may tell us very little about the test performance 'in situ', and can cover a multitude of sins. Lastly, it is equally misleading to give one overall figure, with children's tests, as a measure of the relationship between test scores and intelligence. Though no hard data has been extracted to confirm this in the present study, observation suggests (if it is not intuitively obvious) that the relationship between test scores and intelligence is inversely related to age. Given that a certain minimal degree of intelligence is necessary to master the test situation, and that this remains constant for age groups, the task is likely to be nearer to the intellectual 'ceiling' of a young child than an older child. Thus, when a coefficient of 0.3 is given as a description of the relationship with I.Q., this may mean that whilst there is virtually no relationship for 12 or 13 year olds, there may be a considerable one for 6 or 7 year olds. The person using the tests needs data that are applicable to the sample he is using, rather than some overall measure. All data should be closely related to age therefore, if test users are not to be mislead about the effectiveness and efficiency of the tests in certain situations. It is necessary to make a final point about the present study. All the testing described has taken place in the North East. It is impossible why to state with certainty just how far the findings might be specific to the area, and how far they might be general. Certainly, studies have shown fluctuations in intelligence in different areas, and it does not seem unreasonable to suppose that the present findings might also vary within certain limits if the tests were administered in different geographical regions. Caution should be used, therefore, in interpreting standard scores; though it is unlikely that any of the general principles will be fundamentally altered by location of testing.

Some of the findings from the present study perhaps help to clarify one of the central issues in the psychology of music, namely the 'omnibus' versus the 'atomistic' viewpoints. As so often happens in cases of this type, the truth will probably emerge, eventually, in a position somewhere between the two. This seems very likely, as far as the present test is concerned, since the brief factorial study shows that the tests have considerable specificity, yet all have loadings on substantially one common factor. We might conclude with caution that some evidence has been produced by the present study which suggests that the common factor is a short-term memory factor of a highly specific nature, (i.e. specific to musical materials). There may exist, within this general category, differences in the short-term memory capacities of subjects for different types of musical material, but memory for these different materials is not totally exclusive for each type. Clearly, high scores on one sub-test do not necessarily imply high scores on the others. The 'complexity' of musical ability, which even the 'omnibus' theorists concede, may stem from the variety of musical materials involved in the tests.

It is not proposed to give any detailed account of the role of this type of research in the general sphere of musical education, or education as a whole; nor to give any explanation of the possible value, if any, of this work. The former has been dealt with fully by Wing, Shuter, Bentley, and others, and no purpose is served here by mere repetition. With regard to the value of this type of research, the comments made along the way must be the only guide; no deeper, perhaps pseudo-philosophical discussion, is given. In the ultimate analysis the individual reader must evaluate the work and reach his own conclusions, as he sees fit. Neither is it proposed to discuss the undoubted value of music in therapy (Gaston, T., 1968), in the study of personality (Payne, E., 1967), applications in industry and effects upon industrial performance, (Yoshida, T., 1965) or any general problems concerning the "mystery of music". (Wing, H. and Bentley, A., 1966). All these things are seen as peripheral to the main point of the present study.

It remains finally to look briefly at the position of the present work in the overall picture of musical and auditory research, and to suggest directions for future research.

Work in the field of musical abilities, and work in auditory perception, have often been regarded as quite separate areas of research. There has been very little carry over from the laboratory-experimental studies of hearing, to the less experimental studies of music and musical perception. There are signs in recent years that the more thoroughgoing experimentalists are becoming more and more interested in problems concerned with the perception of musical material. Newman Guttmann, R. J. Ritsma, S. S. Stevens, Merle Lawrence, B. L. Cardozo and many others have performed work on musical perception in a purely experimental setting. The mutual effects of this drawing together can only be beneficial. Test constructors, and musical educationalists can gain insight into certain perceptual processes, and better appreciate the extent of certain abilities, by an examination of more purely scientific data. This must effect their aims, methodology, and selection of material in a beneficial manner; and also help to show what is 'musical truth' and what mere 'musical humbug'. The present work can be regarded mainly as an exercise in psychometrics; and an attempt to modify the type of material used in test batteries in the light of simple psychometric and psychological theory. The result, it is hoped, is firstly a battery of tests that works; and one that may possibly have certain advantages over others in terms of 'culture freeness'. Secondly, the work demonstrates the usefulness and viability of the general approach to measurement; one does not have to use the musicians own criteria when measuring his skill.

On a wider front, certain of the experimental studies have shown problems to exist where the non-experimental worker never suspected them. In particular, some animal experiments have produced results that are very difficult to explain. The work of Dewson has been briefly mentioned in an earlier chapter; he used monkeys in an experiment which showed that the animals showed no preference for (in an operant sense, could not discriminate between) consonance and dissonance. Two pieces of experimental work are particularly interesting. Blackwell and Schlosberg showed, as early as 1943, that white rats showed quite marked octave generalisation in an operant situation. However, D. Allen (1967) recently showed that 'normal' non-musical adult people do NOT show octave generalisation, though musical subjects do. What can we conclude from these pieces of information? Is the 'normal' rat more 'musical' than the normal adult human? Though we may fairly easily explain the difference between the musical and non-musical humans, it is difficult to see why the rat should display such superior performance.

Finally, research of a physiological nature will inevitably take place in these areas, once a sufficient amount of behavioural data has been accumulated. A recent paper by Diana Deutsch outlines certain physiological possibilities for music recognition, and it is only a matter of time before the investigation of the physiology of hearing reaches the same level as investigations of the eye.

In a sense, however, the music educator faces problems that are much broader than those facing the pure experimentalist. He is always confronted by vast variations in individual performance and capacity; there are no nicely balanced control groups for him. Also, he has not the time to sit back and theorise, and in due course of time come up with an answer to a specific question. He is confronted now with the offspring of the country; amidst the clamour he must find the 'right' answers to broad questions which themselves are not precisely formulated. And he must give his answer now. In concluding, it is perhaps fitting to restate the six tenets of 'The Child's Bill of Rights', drawn up by the Music Educators National Conference, in America, 1964.

1.

Every child has the right to full and free opportunity to explore and develop his capacities in the field of music in such ways as will bring him happiness and a sense of well-being; stimulate his imagination and stir his creative activities; and make him feel so responsive that he will cherish and seek to renew the fine feelings induced by music.

2.

As his right, every child shall have the opportunity to experience music with other people, so that his own enjoyment shall be heightened, and he shall be led into greater appreciation of the feelings and aspirations of others.

3.

As his right, every child shall have the opportunity to make music through being guided and instructed in singing, in playing at least one instrument both alone and with others, and, as far as his powers and interests permit, in composing music.

4.

As his right every child shall have the opportunity to grow in musical appreciation, knowledge and skill, through instruction equal to that given in any other subject in all the free public educational programmes that may be offered to children and youths.

5.

As his right, every child shall be given the opportunity to have his interest and power in music explored and developed, to the end that unusual talent may be utilised for the enrichment of the individual and society.

6.

Every child has the right to such teaching as will sensitize, refine, elevate and enlarge not only his appreciation of music, but also his whole effective nature, to the end that the high part such developed feeling may play in raising the stature of mankind may be revealed to him.

APPENDIX ONE.

This section contains brief descriptions of some of the better known musical ability tests. Also included is a table showing the various types of material which have enjoyed wide use, together with a list of workers who have employed them.

The Seashore Measures of Musical Talent.

This battery has undergone two revisions (1939 and 1960) since it first appeared in 1919. There is little difference between the versions of 1939 and 1960, and the 1939 version is the one described here. (In the 1960 version, the title is changed to 'Measures of Musical Talents') The tests are based on the 'atomistic' conception of musical ability, and with the exception of the test of tonal memory all the tonal items are tests of auditory acuity. All material in the test is of a non-musical or quasi-musical nature.

Seashore writes of his tests (Music Educators Journal, October, 1937): "They represent the theory of specific measurements insofar as they conform to the two universal scientific sanctions, on the basis of which they were designed; namely, that (1) the factor under consideration must be isolated in order that we may know exactly what it is that we are measuring, and that (2) the conclusion must be limited to the factors under control.

Each of these six tests purports to measure one of six capacities or abilities for the hearing of musical tones. There is little overlapping in these functions, and their isolation for the purpose of measurement has been criticised only in the case of one".

The six measures in the battery are as follows:-

Pitch.

The subject is presented with a pair of tones, one after the other. He is asked to indicate whether the second tone is higher or lower than the first. Differences, according to Shuter, vary from 2 c.p.s. to 17 c.p.s. The smallest interval (calling for the highest degree of discrimination) is about one cent which represents a one-hundredth part of an equal semitone, or a change of about 1/17th% of a given frequency. Pure tones are used throughout, and there are fifty items.

Time.

The subject is presented with a pair of tones, consecutively. The pitch of the tones remains the same for both presentations, but the duration of the tones is different. The subject must indicate whether the second tone is longer or shorter than the first. (In the 1919 version the subject had to compare two intervals of 'empty' time, as distinct from 'filled' time in the present version; that is, the subject had to compare the duration of two silent periods, the onset and finish of which were indicated by clicks.) Duration differences range from .5 to .03 seconds. There are fifty items.

Intensity (Loudness).

Two consecutive tones, or buzzer noises, are sounded, and the subject has to indicate whether the second one is louder or softer than the first. Range of intensity differences is from 4 decibels down to 0.5 decibels. There are fifty items.

Timbre.

Two complex tones are sounded, consisting of a fundamental and five upper partials. The energy of the third partial is reduced and added to the fourth in steps of 10, 8.5, 7, 5.5, and 4 decibels. The subject is asked to choose the 'best' timbre, which is judged, in the test, to be the one in which the fourth has the lowest intensity. F sharp is used as the standard reference tone, and there are fifty items.

(The above test replaced a test of consonance in an earlier version of the Seashore Battery. In the 1919 version, two intervals were sounded and the subject had to indicate which of the intervals blended better. This type of test material involves certain assumptions about the nature of consonance and dissonance which are open to question. (see page 286)

In the 1960 version, the timbre test is as above but subjects are merely asked to indicate 'same' or 'different', rather than make some type of value judgement with an external criterion. (Shuter, 1968)

Memory -(Tonal Memory).

A tonal sequence, "purposely selected to form no melodic line" (Wing), is played. A second playing follows in which one note is altered, and the subject is asked to pick out which one is different. There are ten items each of length three, four, and five tones; the total therefore being thirty items.

Rhythm.

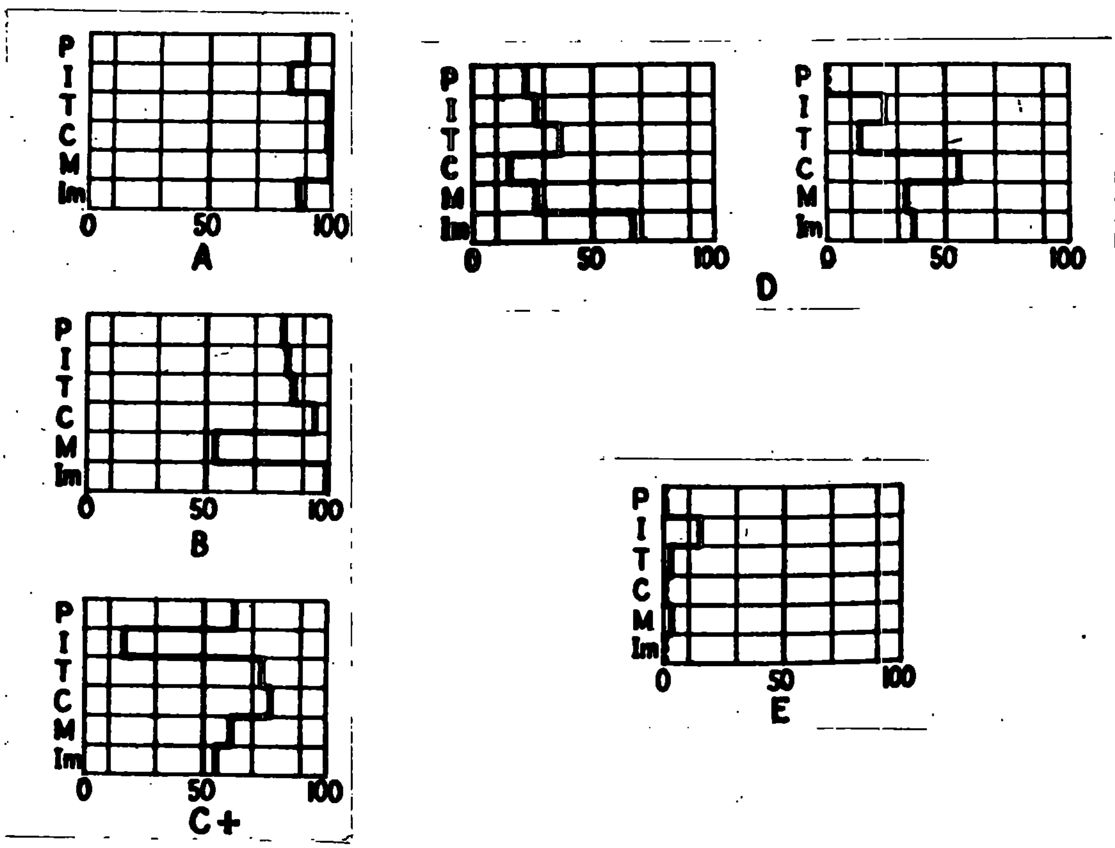
The subject hears two tapped time patterns, one after the other, and is asked to indicate whether the two are the same or different. (1919 version)

In the 1939 version, tones are used in place of the taps. The stimulus thus has the three parameters of loudness, inter-element (temporal) spacing, and duration. Modern workers (Thackray) maintain that all three must be present, and imply that tapping sounds do not in fact have duration in the required sense. It is not thought by the present writer that the taps used by Seashore were particularly inadequate, especially when the alternative is to use tonal material, for which a separate short term memory factor, as distinct from memory for rhythmic material, exists. Total time to administer, about one hour.

The above six tests comprise Seashore's battery, which is available in the form of records. The following are one or two other points of interest about the Seashore battery.

Scoring.

This is based closely on Seashore's conception of musical ability as being 'atomistic', and consists of the construction for each individual of a scoring profile. The profiles obtained were classified according to score, as A (highest), Safe; B Probable; C Possible; D Doubtful; E (lowest), Discouraged. Below are given some sample profiles for an early version of the test.



The profiles sketched above are from Seashore, Psychology of Music, and refer to one of the earliest forms of the test. The horizontal rows are labelled with the first letter of one of the specific abilities, and are the same as for recent versions, with the exception of 'C' for consonance, and 'Im' for imagination. The profiles still serve as an illustration of the scoring method. Of his measures, Seashore writes, "These measures are not in themselves an adequate measure of musical talent as a whole. They are merely a selected battery of measures of specific talents in which a certain degree of capacity is essential to success in music, and in which a certain degree of incapacity is often the basis for failure in music". (Psychology of Music, p.314)

Herbert Wing's Standardised Tests of Musical Intelligence.

Wing first started to work in the field of musical ability testing in 1933. The basis for the present test battery is contained in an M.A. thesis, Tests of Musical Ability in School Children, 1936; and a Ph.D. thesis, Musical Ability and Appreciation, 1941. There has been subsequent revision of individual items since then, but the test has remained basically unchanged in form. All items are presented on a piano, and the material is of a formally musical nature. Wing believed in the 'essential oneness' (Shuter) of musical ability.

Wing writes of his tests, (Manual for Standardised tests of musical Intelligence), "The present set of tests were compiled to correspond to the main characteristics of good music, and are of such a nature that they need no knowledge of musical technicalities for their solution". "The tests can point out possibilities, but cannot guarantee success where interest or the power of persistence is lacking, or in cases where the necessary nervous muscular control does not exist".

There are seven tests in the battery, the last four of which are tests of appreciation. "There are practice items for Tests 1 to 3, but not for Tests 4 to 7. With the later tests (on appreciation) there should be no explanation, e.g. on what is good or bad harmony".

The seven measures in the battery are as follows:-

Chord Analysis.

A chord is played, and subjects are asked to indicate how many tones comprise the chord. The answer sheet provides for a range of answers from 1 up to 6 notes, though in fact no item contains more than 4 notes. In this section, as in all the sections of the battery, subjects are instructed to guess, whenever they are in doubt about an answer. There are four practice items, and twenty testitems.

<u>Pitch Change</u>. (Detecting an alteration of a single note in a repeated chord)

The subject hears two consecutive chords. The second chord may or may not contain one note that is different from the constituents of the first chord; the subject must indicate whether the chords are the same; or if they are different, whether the altered note has moved up or down. There are three practice items, and thirty test items. There are 12 correct 'up' responses, 12 correct 'down' responses, and 6 correct 'same' responses. This means that if a person is unable to do the task, a bias towards guessing 'different' (that is 'up' or 'down') rather than 'same' will probably earn him a far higher score, (i.e. 12 points as against 6).

<u>Memory</u>.

The subject hears a short tune. After a short pause the tune is repeated, and one note is altered. The subject must indicate which note was altered in the second playing. The notes are given numbers to enable the subject to identify them. If the two tunes are the same the subject is explicitly asked to "write S". There is no indication on the answer sheet as to where the "S" should be written, and no provision in the individual items for this. In fact, <u>all the items are different</u>, and the "S" response was included in an attempt to reduce the guessing score. Since the approximate guessing score <u>without</u> the 'same' category is only 5.5 out of 30, the reason for this is not apparent to the writer. (Wing writes, "The marks obtainable by guessing were reduced to one third of the total by including some items in which the two performances were the same", refering to his appreciation tests. This would only be true if the likelihood of any one of the three alternatives being correct was equal, and this accounted for all possible outcomes, i.e. probability of correct guess on any response = $\frac{1}{3}$. This situation does not obtain in Wing's battery)

There are three practice items, and thirty test items. The tunes used vary in length from 3 notes up to 10 notes.

The remaining four tests are tests of appreciation, with the subject having to pick out which rendition is 'best' according to some external criterion.

<u>Rhythmic Accent</u>.

The subject hears two renditions of a tune. Both renditions are identical in tonal terms, but the placing of the accentuated notes may be different. The subject has to indicate whether the two items are the same; if they are not he must indicate which rendition has "the style of playing which you (the subject) think better fits the tune. The material used consists of short extracts from selected "good" music, e.g. Bach chorales, folk songs, etc. There is unequal distribution of correct responses. There are 14 test items. (An earlier version had 20 items in each of the last four tests. These were cut to fourteen for reasons of length, and resultant fatigue.)

Harmony.

As above, except that the two renditions are the same in all respects except harmony. The subject must indicate which performance he thinks is the better; or whether both renditions were the same. There is unequal distribution of correct responses. 14 test items.

Intensity.

As above, except that the two renditions sometimes have the louder and quieter portions in different places, being identical in other respects. The subject must indicate whether the two renditions are the same; of if they are not, state "which style of playing better fits the tune". There is unequal distribution of responses; and again 14 items.

Phrasing.

As above, except that the second playing sometimes has the notes differently grouped. Wing describes this test as follows:-

"The same piece of music is played twice. Sometimes the second playing has the notes differently grouped (different groups of notes may be played with short sharp strokes, or so that they follow on smoothly, etc.). The general effect may be compared to punctuation - that is the use of commas, etc., in ordinary writing. If the two playings are the same, mark 'S'. If they are different choose the style of playing which you think better fits the music. If in doubt, then guess". There is unequal distribution of correct responses. There are 14 items.

The tests are available in the form of a pre-recorded tape, with instructions given verbally, in addition to the written ones on the answer sheet. Duration is about one hour, though the first three items can be used in isolation and last only about twenty minutes. There is probably a fatigue effect towards the end of the full battery. The front page of the answer sheet contains a short questionnaire on general musical attitude and musical experience. On three test tapes used by the present writer there was considerable 'print through', an induction effect which causes each sound on the tape to be followed by several distinct 'echoes'.

<u>Scoring</u>.

One mark is scored for each correct answer, and scores for each section are lumped together into a single total score. The total possible is 136. From the score it is possible to compute the 'Approximate Musical Age', and then the musical quotient or M.Q. Formulae for these have already been given. (page 8) Norms are available for all ages from 8 to 17 years, with scores divided into five categories, generally along the same lines as Seashore. Separate norms are available for tests one, two and three, as these may be used in isolation from the rest of the test. Experiences with young children incline the present writer towards the opinion that the length of the total battery is far in excess of the capacity of the children to concentrate, at a single session.

Of his approach to testing, Wing writes, "It would......appear that at the present stage in music testing it is not possible to name ' a priori' isolated factors which, when added together, make up general musical capacity, and which can be tested for in isolation from music as normally heard. It would therefore seem preferable to approach the problem from an empirical viewpoint by finding out those tests which prove the most efficient as judged by their agreement with the music teachers estimates. This is the method I have adopted".

Arnold Bentley. Measures of Musical Ability.

The Bentley Measures of Musical Ability appeared in 1966, in conjunction with a book entitled 'Musical Ability in Children and its Measurement'. In some degree, the Bentley tests represent a position midway between the extreme 'atomism' of Seashore and the extreme 'unitary ability' position of Wing. Low interest correlations lead him to favour the 'atomistic' view, but his test material is on the whole much less oriented towards peripheral measures than that of Seashore; though the pitch test does involve discrimination of smaller-than-semitone differences.

The four sub-tests in the battery are as follows:-

Pitch Discrimination Test.

In this test the subject is presented with two consecutive tones. He has to judge whether or not the two tones are the same. If not, he must indicate whether the second tone is higher or lower than the first. Concert 'A' (440 c.p.s.) is used as a reference tone, and all pitch differences are deviations from this tone. Pitch differences range from 3 c.p.s. to 26 c.p.s. (Shuter) There are 20 items. Tones are produced on a sine-wave audio-oscillator.

Tonal Memory (Tunes).

The subject hears a little tune of five notes. After a short pause a similar tune is heard. The subject must state whether or not the two tunes are the same; if they are not he must indicate the number of the altered note, in a manner similar to the Wing test of memory. In the tunes, all notes are of equal length, so there is no rhythmic complexity. Although the subject has the option of answering 'same', all the items are in fact different. From a point of view of decision theory this situation is perhaps not ideal. The material is played on a pipe organ, using four-foot and eight-foot flute stops. There are ten items.

Chord Analysis.

A chord is played, and the subject must state how many notes are in the chord. The test is made up entirely of two, three and four-note chords. There is no indication in the instructions as to what possible alternative answers exist; though this situation is the same for all subjects. (But one might expect an increase in the scores of some low scoring subjects if possible alternatives were made clear) Material is again played on a pipe organ, using an eight-foot open diapason stop. This does not produce a sine-wave.

There are twenty items, (Shuter gives the number incorrectly as 10) and each chord has a duration of three seconds.

Rhythmic Memory.

This consists of a series of 'paired-comparisons' after the fashion of the tonal memory test. The subject hears a rhythmic pattern, based on an underlying metre. (All items are in simple quadruple time) After a short pause he hears another time pattern. He must state whether the two patterns are the same; if they are different, he must give the number of the beat (not the number of the note as in the tonal memory test) on which the change occurs. Each half of each item is one measure (bar) long. In this test, 'same' items do in fact occur in equal proportions to other correct answers.

There are ten items; means of presentation is via a pipe organ using eight-foot small diapason and two-foot fifteenth stops. Pitch is kept constant throughout the tests. There are ten items.

Scoring.

One mark is awarded for each correct answer, the total possible being 60 marks. Scores are again placed in five categories, A, B, C, D, and E, in groups of 10, 20, 40, 20, and 10 percent according to relative score.

The tests are available in record form. The actual tests last only about twenty minutes, and total administration time rarely exceeds a half hour. All instructions are given verbally on the record.

Bentley writes, "It is interesting to note that, in spite of their criticisms of the analytic approach, when the adherents of the 'music is a unity' school of thought themselves attempt to measure musical ability the very complexity of this ability compels them to use separate tests of the different factors in it.

It has to be accepted that the measurement of musical ability has not yet progressed beyond a rather rudimentary and unsatisfactory stage. Nor could it be otherwise whilst there exists no agreement on what musical ability is. We may be able to recognise it, or think we can, but we cannot as yet define it".

The above is a brief description of the three tests which are judged by the present writer as being perhaps the most representative of the existing batteries, and also they have had fairly widespread use in this country. A considerable number of other tests exists, however, which it is not proposed to discuss here. References for these may be found at the back. There is a certain amount of redundancy in much of the test material, and below is given a table showing the broad types of test material used by different writers, wherever overlap occurs. The table is not exhaustive, but points out only <u>major</u> similarities in test procedure.

AUTHORS USING SIMILAR TEST MATERIAL.

Pitch Tests.	1. Mainwaring. 2. Seashore. 3. Kwalwasser-Dykema. 4. Bentley. 5. Wing (but with masking tones)
Chord Analysis Tests.	1. Wing. 2. Bentley.
Memory Tests (a)	1. Wing. 2. Bentley. 3. Seashore. 4. Kwalwasser-Dykema. 5. Gaston. 6. Whistler and Thorpe.
Memory Tests (b)	1. Drake. 2. Kwalwasser.
Interval Tests.	1. Madison. 2. Lundin.
Rhythm Tests (a)	1. Seashore. 2. Bentley. 3. Kwalwasser-Dykema. 4. Whistler and Thorpe. 5. Thackray.
Rhythm Tests (b)	1. Thackray. 2. Drake.
Phrasing.	1. Gordon. 2. Wing.

KEY.

Pitch Tests. Paired comparison of two tones, requiring 'same/different' and/or 'up/down discrimination.

Chord Analysis Tests. "How many notes in this chord?"

Memory Tests (a). Paired comparison of tonal sequences, requiring 'same/different' discrimination, with possible 'location of changed note' required.

Memory Tests (b). Paired comparison of tonal sequences, requiring 'nature of change' discrimination.

<u>Interval Tests.</u> Comparison of simultaneous (harmonic) intervals, requiring 'same/different' discrimination.

<u>Rhythm Tests (a).</u> Paired comparison of two rhythmic sequences or patterns requiring 'same/different' discrimination, with possible location of changed element required.

<u>Rhythm Tests (b).</u> The subject must continue to count a steady beat established initially by a metronome, through a period of silence, until the signal to stop is given. If the subject continues to count steadily and at the proper speed, he will arrive at the correct number by the time the signal to stop is given.

<u>Phrasing.</u> Paired comparison of two fragments of music. The subject must indicate which performance is better in terms of musical phrasing.

BIBLIOGRAPHY.

AARONSON, D., 1967. Temporal Factors in Perception and Short-Term Memory. Psych. Bull. 67, 2, 130 - 144.

ALLEN, D., 1967. Octave discriminability of musical and non-musical subjects. Psychonomic Science 1967, 7, 421 - 22.

ANASTASI, A., 1965. Individual Differences. Perspectives in Psychology Series, John Wiley and Sons, Inc.

ANASTASI, A., 1961. Psychological Testing (2nd Ed.) The Macmillan Company, N.Y.

A.P.A., 1955. Standards for Educational and Psychological Tests and Manuals. Report. The American Psychological Association.

BACHEM, A., 1954. Time factors in pitch determination. J. Ac. Soc. Am., 26, 5, Sept., 1954.

BACHEM, A., 1948. Note on Neu's review of the literature on absolute pitch. Psych. Bulletin, 45, 161 - 162.

BARNETTE, W. L., Jr. (Ed.)., 1964. Readings in Psychological Tests and Measurements. The Dorsey Press Inc,, Illinois.

Von BEKESY, 1960. Experiments in Hearing. McGraw Hill Book Company Inc.

BELL, F., 1928. Hugh Lowthian Bell. A Record and some Impressions. 1878 - 1926. Middlesborough:- Wllm. Appleyard and Sons Ltd. Private circulation only.

BENTLEY, A., 1966. Musical Ability in Children and its Measurement. Harrap and Co. Ltd.

BERGEL, E. E., 1962. Social Stratification. McGraw Hill Book Company, Inc.

BERNSTEIN, L. 'I want to be in America' (song) from "West Side Story".

BERRIMAN, J. M., 1968. (Supervised by J. B. Davies). The Differential Effect of Musical and Non-Musical Interference on musically "trained" and "untrained" persons. Unpublished dissertation. Durham University Psychology Department Library.

BIJOU, S. W. and BAER, D. M., 1960. The Laboratory Experimental Study of Child Behaviour. From Handbook of Research Methods in Child Behaviour. Ed. Mussen, P. H. p. 140.

BLACKWELL, H. R. and SCHLOSBERG, H., 1943. Octave generalisation, pitch discrimination and loudness thresholds in the white rat. J. Exp. Psy., 1943, 33, 407 - 419.

BOGGS, L. P., 1907. Studies in Absolute Pitch. Am. J. Psy. 1907, 18, 194 - 205.

BORING, E. G., 1923. Intelligence as the tests test it. New Republic. 1923, 34, 35 - 37.

BREHMER, F., 1925. Melodieauffassung u. Melodische Begabung des Kindes. Leipsig, 1925. (from Teplov)

BROADBENT, D. E., 1958. Perception and Communication. Pergamon Press.

BROWN, A. W., 1928. The reliability and validity of the Seashore tests of musical talent. J. Appl. Psychol., 12, 1928.

BURT, C., 1947. Mental and Scholastic Tests. (2nd Ed.) Staples Press Ltd.

CARDOZO, B. L., 1967. Ohm's Law and Masking. I.P.O. Annual Progress Report. No. 2, pp. 59 - 64.

CARDOZO, B. L. and van NOORDEN, 1968. Imperfect periodicity in Bowed String. I.P.O. Annual Progress Report, No. 3., 23 - 29.

CERASO, J., 1967. The Interference Theory of Fogetting. Scientific American, 217, 4, 117 - 121.

CHERRY, C., 1966. On Human Communication (2nd Ed.), Chapter 5, On the Statistical Theory of Communication. The M.I.T. Press.

CLACK, T. D., 1967. Aural Harmonics: The masking of a 2000-Hz Tone by a Sufficient 1000-Hz Fundamental. J. Acous. Soc. Am., 42, 4, pp. 751 - 758.

CLEALL, C., 1968. The Natural Pitch of the Human Voice. Paper presented at Annual B.P.S. Conference (Ed. section) September, 1968.

COHEN, A., 1961. A further investigation of the effects of intensity upon the pitch of pure tones. J. Acous. Soc. Am., 32, 1308 - 1319.

CONRAD, R., 1959. Errors of Immediate Memory. Brit. J. Psych., 50, pp. 349 - 359.

CORCORAN, D. W. J. and WEENING, D. L., 1967. Redundance Effects in Short-Term Memory. The Quarterly J. of Exp. Psy. XIX 4, Nov., 1967, 309 - 318.

CORNSWEET, T. N., 1963. The Design of Electric Circuits in the Behavioural Sciences. John Wiley and Sons, Inc.

CRONBACH, L. J. and MEEHL, P. E., 1967. Construct Validity in Psychological Tests. From Problems in Human Assessment, Jackson, D. N. and Messick, S. McGraw Hill, 1967. Chapter 4.

CURETON, E. E., 1950. Validity, Reliability and Baloney. Ed. and Psych. Measurement, 10, pp. 94 - 96.

CURWEN, J. S. No date given. Psychology applied to Music Teaching. Curwen edition 8307. London:- J. Curwen and Sons Ltd.

DAVIDOFF, M. D. and GOHEEN, H. W., 1953. A table for the rapid determination of the tetrachoric correlation coefficient. Psychometrika 18, 115 - 121.

DAVIES, J. B., 1965. An Examination of Musical I.Q. and Musical Preference. Unpublished dissertation. University of Durham Psychology Dept. Library.

DAVIES, J. B. and BILLINGS, C. A., 1967. An Investigation of Musical Ability. Unpublished Study. University of Durham Psychology Dept.

DAVIES, J. B. and JACKSON, C., 1967. Social Environment and Musical Attitudes. Unpublished study. University of Durham Psychology Dept.

DEUTSCH, D., 1969. Music Recognition. Psych. Review. 76, 3, 300 - 307.

DEWSON, J., 1969. Personal Communication.

DINGLE, H., 1960. Basic Problems of Measurement. Scientific American v. 202, No. 6. from Barnette, W. L.

DOBBS, J. P. B., 1966. The Slow Learner and Music. Oxford Univ., Press, 1966.

DRAKE, R. M., 1933. Four New Tests of Musical Talent. J. Appl. Psych. 17, 136 - 147.

DRAKE, R. M., 1931. Tests of Musical Talent. Ph.D. Thesis, London University. (from Shuter)

EBBINGHAUS, 1897. Development of Completion Tests. Breslau.

EBEL, R. L., 1961. Must All Tests be Valid? American Psychologist. 16, 640 - 647.

EGAN, J. P., 1958. Recognition Memory and the Operating Characteristic. Indiana University. Technical Note AFCRC-TN-58-51, 1958. (from Green and Swets)

ELLIS, A. J., 1954. Footnote to page 11. The Sensations of Tone.

FAULDS, B., 1959. The Perception of Pitch in Music. Princeton University; Princeton, N. J. (from Shuter)

FLETCHER, H., 1953. Speech and Hearing in Communication. D. van Norstrand Co. Ltd., Princeton.

FLETCHER, H. and MUNSON, W. A., 1933. Loudness, its definition, measurement, and calculation. J. Acous. Soc. Am., 1933, 5, 82 - 108.

FRANKLIN, E., 1956. Tonality as a basis for the study of Musical Talent. Göteberg, Sweden: Gumpert.

GARDNER, S., 1939. School for Violin Study Based on Harmonic Thinking. Carl Fischer Inc. New York.

GASTON, T., 1968. Music in Therapy. McMillan, N.Y.

GESELL, A., and ILG, F., 1946. The child from five to ten. London: Hamilton.

GETZELS, J. W. and JACKSON, P.W., 1961. A study of the Sources of Highly Intelligent and of Highly Creative Adolescents. Am. Soc. Rev., 1961.

GETZELS, J. W. and JACKSON, P. W., 1962. Creativity and Intelligence. Wiley, 1962.

GIBSON, E. J. and OLUM, V., 1960. Experimental Methods of Studying Perception in Children. From Handbook of Research Methods in Child Behaviour. Ed. Mussen, P. H. Note on pitch discrimination, p. 333.

GREEN, D. M. and SWETS, J. A., 1966. Signal Detection Theory and Psychophysics. John Wiley and Sons, Inc.

GUILFORD, J. P., 1967. The Nature of Human Intelligence. McGraw Hill Book Company.

GUILFORD, J. P., 1959. The Structure of Intellect. Rev. Ed. Res. 29, 26 - 30.

GUILFORD, J. P., 1952. When Not to Factor Analyze. Psych. Bull, 1952. (from Jackson and Messick)

GUTTMAN, N., 1965. Binaural Interaction of Three Clicks. J. Ac. Soc. Am., 37, 1, 145 - 150.

GUTTMAN, N. and JULESZ, B., 1963. Lower Limits of Auditory Periodicity Analysis. J. Acous. Soc. Am., 35, 4, 610.

GUTTMAN, G. and PRUZANSKY, S., 1962. Lower Limits of Pitch and Musical Pitch. J. of Speech and Hearing Res. 5, 3, 207 - 214.

HARGREAVES, D., 1969. (Supervised by J. B. Davies) Some problems in the Measurement of Divergent Musical Ability. Unpublished dissertation. University of Durham Psychology Department Library.

HEINLEIN, C. P., 1929. A New Method of Studying the Rhythmic Responses of Children. J. Genet. Psycho., 36, 205 - 228.

HEINLEIN, C. P., 1925. An Experimental Study of the Seashore Consonance Test. J. Exp. Psychol. VIII, 408.

HELMHOLTZ, L. F. H., 1877. On the Sensations of Tone. 2nd English Edition, 1954. Dover Publications Inc. New York.

HENKIN, R. I., 1957. Re-evaluation of a factorial study of the components of music. J. Psychol., 43, 301 - 306.

HEVNER, K. and LANDSBURY, J., 1935. Oregon Musical Discrimination Tests. Chicago; C. H. Stoelting Co.

HICKMAN, A., 1969. Musical Imaging and Concept Formation. Notes on paper read at Conference on Music in Education, Reading University.

HICKMAN, A., 1968. Musical Imaging and Concept Formation in School Children. Unpublished Ph.D. Thesis, University of Manchester, 1968.

HICKMAN, A. T., 1968. Electronic Apparatus for Music Research. Music Education Research Papers, No. 3. Novello and Co. Ltd., London.

HOLMSTROM, L.G., 1963. Musicality and Prognosis. Uppsala: Almqvist and Wiksells.

HOWES, D. and OSGOOD, C. E., 1954. On the Combination of Associative Probabilities in Linguistic Contexts. Am. J. Psy., 67, 241 - 58.

HOLWAY, A. H. and BORING, E. G., 1940. Am. J. Psy., 53, 109, 537.

HULL, C. L., 1929. Aptitude Testing. London. Harrap, 1929.

JACKSON, D. N. and MESSICK, S. (Ed.), 1967. Problems in Human Assessment. McGraw Hill Book Company.

JEFFREY, W. E., 1961. Simultaneous v. successive stimulus presentation. Child development 32, p. 305.

KAHL, J., 1953. Educational and Occupational Aspirations of "Common Man's" Boys. Harvard Ed. Rev. 23, 1953.

KARLIN, J. E., 1941. Music Ability. Psychometrika, 6, 61 - 65.

KOHLER, W., 1910. Akustische Untersuckungen. Zeits. Psychol., 54, 1910; 58, 1911; 72, 1915. (from Teplov)

KWALWASSER, J., 1927. Kwalwasser Test of Musical Information and Appreciation. Iowa City, Bureau of Educ. Research, University of Iowa.

KWALWASSER, J., 1927. Tests and Measurements in Music. Boston: Birchard, 1927.

KWALWASSER, J. and DYKEMA, P. W., 1930. Kwalwasser-Dykema Tests, New York: Carl Fischer, 1930.

LAMP, C. J. and KEYS, N., 1935. Can Aptitude for Specific Musical Instruments be predicted? Amer. J. Ed. Psych. XXVI, 1935, 587 - 596.

LAWRENCE, M., 1968. Audition. Annual Rev. of Psy. 19, 114.

LORENTE, M., 1965. Physical Basis of Musical Tonality and Consonance. 5th Congres International d'Acoustique Liege, Sept., 1965. M 64.

LORGE, I. and THORNDIKE, R. L., 1957. Procedures for Establishing Norms. From Technical Manual, Lorge-Thorndike Intelligence Tests, 1957. The Houghton-Mifflin Publishing Co.

LOWERY, H., 1926. Cadence and Phrase Tests of Musical Talent. Brit. J. Psy. XVII, 1926, p. 111.

LOWERY, H., 1929. Musical Memory. Brit. J. Psy. XIX, 1929, p. 397.

LOWERY, H., 1932. Estimation of Musical Capacity. Proceeds of the Manchester Lit. and Phil. Society. No. 6, 76, p. 53.

LUNDIN, R. W., 1953. An Objective Psychology of Music. New York: Ronald Press.

LUNDIN, R. W., 1949. The Development and Validation of a set of musical ability tests. Psychol. Monogr., 63, 305, 1 - 20.

MACKINNON, D. W., 1962. The Nature and Nurture of Creative Talent. American Psychologist, 1962. (see Jackson and Messick)

MADISON, T. H., 1942. Interval Discrimination as a Measure of Musical Aptitude. Arch. Psychol. 268, p. 100.

McCARTHY, D. A., 1930. Study of the Seashore measures of musical talent. J. Appl. Psychol., 14, 1930. (from Teplov)

McLEISH, J., 1968. The factor of musical cognition in Wing's and Seashore's Tests. Music Education Research Papers, No. 2. Novello and Company Ltd.

McKEE, J. P. and RILEY, D. A., 1962. Auditory Transposition in six-year old children. Child Development 33, p. 469.

McLEISH, J., 1950. The Validation of Seashore's Measures of musical talent by factorial methods. Brit. J. Stat, Psychol., iii, 129 - 40.

MEDNICK, S. A., 1962. Associative Basis of the Creative Process. Psych. Rev. 1962, 69, 220.

MEEHL, P. E. and ROSEN, A., 1955. Antecedent Probability and the Efficiency of Psychometric Signs, Patterns, or Cutting Scores. Psych. Bull. 1955. (see Jackson and Messick)

MEISSNER, H., 1914. Beitrag zur Entwicklung des Musikalischen Sinnes Beim Kinde Während des Schulpflichtigen Alters. Berlin, 1914 (from Teplov)

MEYER, L. B., 1956. Emotion and Meaning in Music. University of Chicago Press, 1956.

MILLER, G. A., 1951. Language and Communication. McGraw Hill Book Company Inc.

MILLER, G. A. and SELFRIDGE, J. A., 1953. Verbal Context and the Recall of Meaningful Material. Am. J. of Psy. 63, 176 - 85.

MOLES, A., 1968. Information Theory and Esthetic Perception. Illini Books. University of Illinois Press.

MOORHEAD, G. E. and POND, D., 1941 - 1942. The Music of Young Children. Pillsbury Foundation. Stud. (from Shuter)

MULL, H. K., 1925. The Acquisition of Absolute Pitch. Am. J. Psy. XXXVI, 1925, 469.

MULLIGAN, R., 1952. Socio-economic Background and College Enrollment. American Sociological Review, 16, 1951.

MURDOCK, B. B., Jr., 1965. Signal-detection Theory and Short-Term Memory. J. Exp. Psychol., 70, 443 - 447.

MURSELL, J. L., 1937. Music Educators Journal, 1937. (from Seashore, Psychology of Music))

MURSELL, J. L., 1937. The Psychology of Music. New York: Norton.

MYERS, C. S., 1904. A Study of Rhythm in Primitive Peoples. Brit. J. Psy., 1904., 1, p. 397.

NEU, D. M., 1948. Absolute Pitch - a reply to Bachem. Psych. Bulletin, 1948, 45, 534 - 535.

NEU, D. M., 1947. A Critical Review of the Literature on Absolute Pitch. Psych. Bulletin, 1947, 44, 249 - 266.

NICKSON, N., 1966. Education Through Music. University of Queensland Inaugural Lectures. University of Queensland Press.

NOBLE, C.E., 1964. The Psychology of cornet and trumpet playing. Mountain Press, Missoula, Mont.

NORMAN, D. A. and WICKELGREN, W. A., 1965. Short-Term Recognition Memory for single digits and pairs of digits. J. Exp. Psychol., 1965, 70, 479 - 489.

OLSON, H. F., 1967. Music, Physics and Engineering. (second edition) Dover Publications, Inc. N.Y.

PARKS, T.E., 1966. Signal-Detectability Theory of Recognition-Memory Performance. Psychol. Rev. 73, 44 - 58.

PAYNE, E., 1967. Musical Taste and Personality. Br. J. Psy. 58, 1 and 2, 133 - 138.

PERRY, N. C., KETTNER, N. W., HERTZKA, A. F., and BOUVIER, E. A., 1953. Estimating the Tetrachoric Correlation Coefficient via 1. a Cosine-Pi Table and 2. correction graphs for non-median dichotomization. Studies of Aptitudes of High-level Personnel, Technical Memorandum No. 2., Dept. of Psy., University of Southern California, Los Angeles.

PHILLIPS, B. N. and WEATHERS, G., 1958. Analysis of Errors Made in Scoring Standardized Tests. From Educational and Psychological Measurement, 1958. G. Frederic Kuder.

PIAGET, J., 1947. The Psychology of Intelligence. (fifth impression) International library of Psychology, Philosophy, and Scientific Method. Routledge and Kegan Paul Ltd.

PIAGET, J., 1941. The Child's Conception of Number. International Library of Psychology, Philosophy and Scientific Method. Routledge and Kegan Paul Ltd.

POLLACK, I., GALANTER, E., and NORMAN, D., 1964. An Efficient Non-Parametric Analysis of Recognition Memory. Psychonomic Science, 1, 327 - 328.

POMFRET, V., 1969. (Supervised by J. B. Davies) An Investigation of Rhythmic Ability in Young Children, and its relationship to Short-Term Memory. Unpublished Dissertation. University of Durham, Psychology Dept. Library.

QUASTLER, H., 1956. Studies of human channel capacity. From Cherry, 3rd London Symposium on information theory, 1955., pp. 363 - 369. Butterworth's Scientific Publications.

RAINBOW, E. L., 1965. A Pilot Study to investigate the Constructs of Musical Aptitude. J. Res. Mus. Ed., 13, 3 - 14.

REVESZ, G., 1925. The Psychology of a Musical Prodigy. New York: Harcourt Brace & Co.

RITSMA, R. J., 1968. Dominance: A Unifying Concept in Pitch Perception. I.P.O. Annual Progress Report, No. 3., 15 - 22.

RITSMA, R. J., 1967. Frequencies dominant in the Perception of Pitch of Complex Sounds. I.P.O. Annual Progress Report, No. 2., pp. 5 - 10.

RITSMA, R. J., 1966. The Octave Deafness of the Human Ear. I.P.O. Annual Progress Report No. 1, 15 - 17.

RITSMA, R. J., 1966. The Pitch of Sinusoids and Complex Signals as Affected by Masking Noise. I.P.O. Annual Progress Report, No. 1., 27 - 28.

RITSMA, R. J., 1965. Pitch discrimination and frequency discrimination. 5th Congres International d'Acoustique Liege B22.

RITSMA, R. J., 1965. Timbre and pitch. Intern. Audiol. 7, p. 67.

RITSMA, R. J. and ENGEL, F. L., 1964. Pitch of Frequency-Modulated Signals. Journal Ac. Soc. Am. 36, 1637 - 1644.

SCHOUTEN, J. F., 1968. The perception of Timbre. I.P.O. Annual Progress Report. No. 3., pp. 32 - 35.

SEASHORE, C. E., 1938. Psychology of Music. McGraw Hill Book Co., Inc.

SEASHORE, C. E., 1937. Two Types of Attitude toward the Evaluation of Musical Talent. Music Educators Journal, December, 1937. Reproduced in Appendix, Psychology of Music.

SEASHORE, C. E., 1919. The Psychology of Musical Talent. Newark, N. J. Silver, Burdett and Co.

SEMEONOFF, B., 1940. Further developments in a new approach to the testing of musical ability, with special reference to groups of secondary school children. Brit. J. Psy., 31, 145 - 61 (1940 b.)

SEMEONOFF, B., 1940. A New Approach to the Testing of Musical Ability. Brit. J. Psy., 30, 326 - 40. (1940 a.)

SHUTER, R., 1968. The Psychology of Musical Ability. Methuen and Co. Ltd.

SLAWSON, A. W., 1968. Vowel quality and Musical Timbre as Functions of Spectrum Envelope and Fundamental Frequency. J. Acous. Soc. Am., 43, 1, 87 - 101.

SPEARMAN, C., 1927. The Abilities of Man, their Nature and Measurement. London: Macmillan and Co.

SPEARMAN, C. and JONES, L. W., 1950. Human Ability. London: Macmillan and Co. Ltd.

STERN, W., 1927. Psychologie der Frühen Kindheit, 4. Aufl., 1927. (from Teplov)

STEVENS, S. S. (Ed.), 1951. Handbook of Experimental Psychology. Chapman and Hall, Ltd., London.

STEVENS, S. S., 1935. The relation of pitch to intensity. J. Acous. Soc. Amer. 1935, 6, 150 - 154.

STEVENS, S. S. and DAVIS, H., 1947. Hearing:- Its Psychology and Physiology. New York: John Wiley and Sons Inc.

STEVENS, S. S. and VOLKMANN, J., 1961. The Relation of Pitch to Frequency: a revised scale. Amer. J. Psy., 53, 329 - 353.

STUMPF, C., 1914. Ueber neuere Untersuchungen zur Tonlehre Ber. VI, Kong. exper. Psychol., 1914 (from Teplov)

TANNER, PATTON and ATKINSON, 1966. The Effect of Signal Intensity on Comparative Judgements of Auditory Durations. Psychonomic Sci., 4, 10.

TAYLOR, D. E., and BORING, E. G., 1942. Am. J. of Psy., 1942, 55, 189.

TEPLOV, B. M., 1966. Psychologie des Aptitudes Musicales. Presses Universitaires de France.

TERMAN, L. M., 1954. The Discovery and Encouragement of Exceptional Talent. Walter V. Bingham Memorial Lecture, University of California, Berkeley. March 25th, 1954.

THACKRAY, R., 1969. An Investigation into Rhythmic Abilities. Music Education Research Papers, No. 4., Novello and Co. Ltd.

THACKRAY, R., 1968. Further Investigations into Rhythmic Abilities. Individual paper, given at Annual Conference of B.P.S. (Education Section), September, 1968.

THORNDIKE, R. L., 1949. The Analysis and Selection of Test Items. From Personnel Selection. John Wiley & Sons, Inc.

TROTTER, J. R., 1967. The Psychophysics of Melodic Interval:- Definitions, techniques, theory, and problems. Australian J. of Psy. 19, 1, 13 - 25.

TURNBULL, W. M., 1944. Pitch discrimination as a function of tonal duration. J. Exp. Psychol. 34, 302 - 316.

VALENTINE, C. W., 1962. The Experimental Psychology of Beauty. London: Methuen.

VALENTINE, C. W., 1913 - 14. The Aesthetic Appreciation of Musical Intervals among School Children and Adults. Brit. J. Psych. VI (1913), 190. VII (1914), 108.

VERNON, P. E., 1969. Intelligence and Cultural Environment. Methuens Manuals of Modern Psychology. Methuen and Co. Ltd.

VERNON, P. E., 1950. The Structure of Human Abilities. London: Methuen.

VERNON, P. E., 1948. Indices of Item Consistency and Validity. Brit. J. Psy. (Stat. Section) 1948, 1, 152 - 66.

WALLACH, M. A. and KOGAN, N., 1965. Modes of thinking in Young Children. Holt. Rinehart and Winston, 1965.

WARD, W. D., 1954. Subjective Musical Pitch. J. Accous. Soc. Am., 26, 369 - 380.

WEDDELL, C. H., 1941. A Study of Absolute Pitch. Psych. Bull. 38, 547 - 48.

WHITFIELD, I. C., 1967. Coding in the Auditory Nervous System. Nature, Vol. 213, No. 5078, 756 - 760.

WICKELGREN, W. A., 1966. Consolidation and retroactive interference in short-term memory for pitch. J. Exp. Psych. 72, 2, 250 - 259.

WICKS, B., 1968. The Poulton Experiments in Creative Music. Paper given at Annual conference of B.P.S. (Education Section) Didsbury, September, 1968.

WING, H. D., 1962. A Revision of the Wing Musical Aptitude Test. The Journal of Research in Music Education. Vol. x, No. 1, Spring, 1962.

WATSON, L. A., and TOLAN, T., 1949. Hearing Tests and Hearing Instruments. The Williams and Wilkins Co., Baltimore.

WHISTLER, H. S. and THORPE, L. P., 1950. Musical Aptitude Test. Los Angeles: Calif. Test Bureau. (from Shuter)

WING, H. D., 1948. Tests of Musical Ability and Appreciation. Monograph Supplement XXVII. The British Journal of Psychology. Cambridge University Press.

WING, H. D., 1941. A factorial study of Musical Tests. British Journal of Psychology XXXI, p. 341.

WING, H. D., 1936. Tests of Musical Ability in School Children. Unpublished Thesis, University of London Library.

WING, H. D. and BENTLEY, A., 1966. The Mystery of Music. New Scientist 32, 520, p. 306.

WOODWORTH, R. S. and SCHLOSBERG, H., 1963. Experimental Psychology (3rd ed.) Holt, Rinehart and Winston, Inc.

YOSHIDA, T., 1965. An Investigation on How Industrial Music in Japan goes on. 5th International Congress on Acoustics. Liege, 1965, M56.

SPECIAL APPENDICES.

APPENDIX ONE.

THE 'A' SCALE.

Part One: Melody.

Practice Items:

1.	E D C	C E♭' B♭, G♯ F♯' G
2.	G A B C'	C C' G A B C' G
3.	F A C'	A G F A C' F♯

Test Items:

1.	B A G	D' F' E' C' E' E
2.	B B♭ B C♯	G' F' A' E' B D' F♯ G
3.	E♭ B♭ E♭'	E♭ B♭ E♭' D' E♭' G'
4.	F' E' D' C'	C♯' C' D' F' B♭ D' B♭ D'
5.	B♭ A G	F' D' B♭ A G B
6.	E' G' F♯' E'	A E' G' G F' B F♯ F♯'
7.	F F♯ G D	A B F F♯ G D E' C'
8.	D B G	F♯' F' D B G A
9.	B A F♯ D	F' D' E' B♭ E♭ G A♭ C♯
10.	G F♯ B	B A G F B F♯
11.	C' D' G C'	G C' E' C' D' G C' E'
12.	F E♭ A♭ G	B F E' C' F A♭ G C♯'
13.	C' D' G A	C' E' D' G A C' G A
14.	B, D B G	B♭ A B, D B G E♭ F
15.	G D' E' F♯	C' G D' E' F♯ A C♯' F♯

Letters indicate tunes comprising the test items in alphabetical notation.

Part Two: Pitch.

Practice Items:

1. 580 2. 970 3. 438

Test Items:

1.	790	9.	1260
2.	990	10.	570
3.	720	11.	1060
4.	600	12.	790
5.	990	13.	740
6.	450	14.	600
7.	390	15.	350
8.	560		

Numbers indicate frequencies of stimulus tones in cycles per second.

Part Three: Auditory Transposition.

Practice Items:

1. DEb FF# DE'
2. CC' DEb CC#
3. EEb' FF# DC#'

Test Items:

1. CG FC' ABb
2. EbC#' CBb D,F#'
3. GA DE' FG
4. AF#' DA GF'
5. A,G# DC#' EG
6. GC' DG# AC#'
7. A,A EF# DD'
8. EG# BD#' CD'
9. GF#' DE ED#'
10. DD EE FC'
11. BbG# CF DE'
12. DF# C'D' CE
13. AC' DA EbF#
14. BbG FG CE'
15. FB CD EE'

First column gives stimulus interval.

Second column gives first comparison interval.

Last column gives second comparison interval.

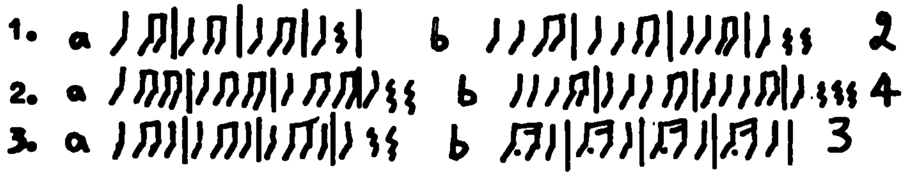
Part Four: Rhythm.

Items comprising the rhythm test are given overleaf. The two comparison rhythms are labelled 'a' and 'b' respectively. The number at the end of each pair indicates whether the preceding metre was duple, triple, quadruple, or quintuple.

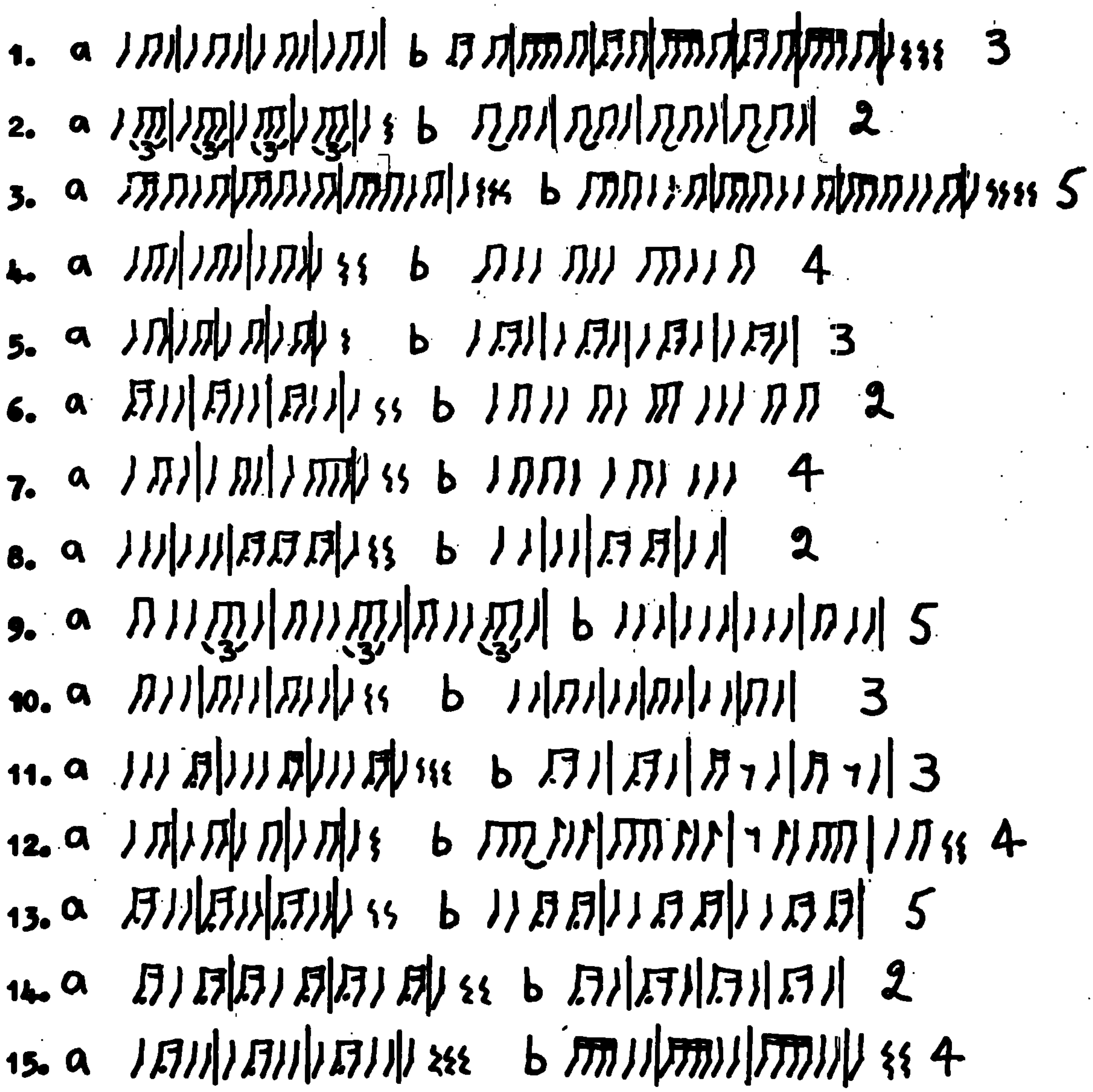
The items comprising form 'B' of the test are laid out in a similar manner to the above version, but on most parts of the 'B' form there are fewer comparison items. (see text)

Part Four: Rhythm.

Practice Items.



Test Items.



THE 'B' SCALE.

Part One: Melody.

Practice Items:

1.	F A G	F A G C
2.	D E F♯	D' A C♯' D E
3.	D F A	D F A D' D

Test Items:

1.	B G D	D' F' E' C'
2.	B♭ B C♯	B D' F♯ G
3.	E♭ B♭ E♭'	E♭ B♭ E♭' D'
4.	F' E' D'	C♯' C' D' F' B♭
5.	B♭ A G	D' B♭ A G B
6.	G F♯ E	E G' G F B
7.	F F♯ A	F F♯ A D E'
8.	G G♯ E♭'	F' E' G G♯ E♭' G'
9.	A F♯ D	F♯ E♭ G G♯
10.	G F♯ B	B A G F
11.	C' D' G C'	G C' E' C' D' G C' E'
12.	F E♭ G♯ G	B F A C' F G♯ G C♯'
13.	C' D' G A	C' E' D' G A C' G A
14.	B, D B G	B♭ A B, D B G E♭ F
15.	G D' E' F♯	C, A G D' E' F♯ C♯' F♯

Part Two: Pitch.

Practice Items:

1. 610
2. 387
3. 786

Test Items:

1.	235	9.	175
2.	400	10.	525
3.	255	11.	350
4.	650	12.	220
5.	390	13.	800
6.	150	14.	455
7.	420	15.	210
8.	310		

Part Three: Auditory Transposition.

Practice Items:

1.	E	E'	F	G
2.	D	E	F	F♯'
3.	G	A	E♭	F

Test Items:

1.	C	E	D	D'	9.	D	A	E	F♯
2.	C	C'	G	A	10.	C	F♯	D	B♭
3.	E	F♯'	E	G	11.	C	E	D	F♯
4.	D	F♯	F	A	12.	D	E'	F	C'
5.	G	B	D	C'	13.	D	B♭	F	A
6.	C	C'	D	D'	14.	C	C♯	B	C'
7.	E	G♯	F	D'	15.	E	C♯'	C	A
8.	E	G	G	E♭'					

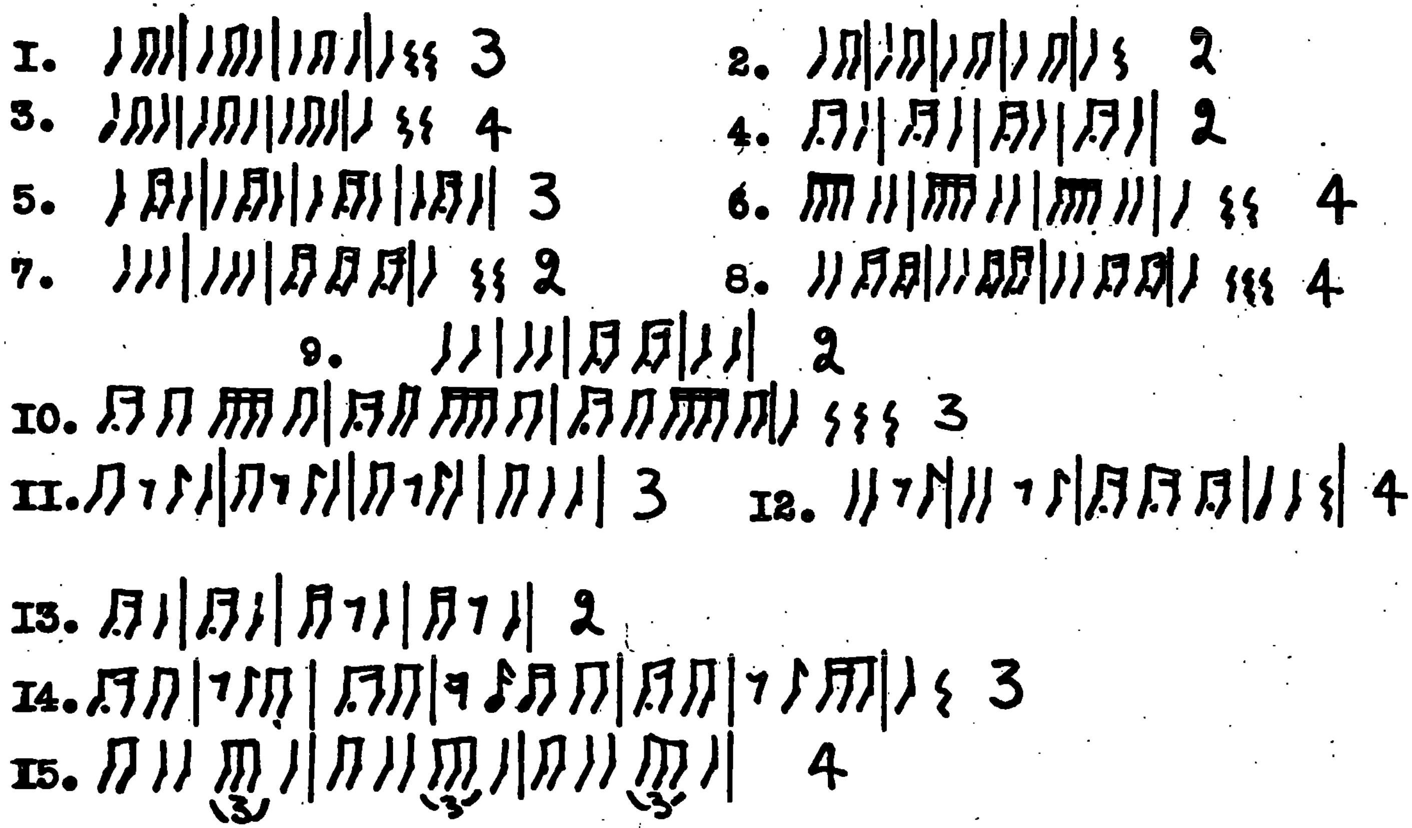
Part Four: Rhythm.

Items comprising the rhythm test in form 'B' appear overleaf. Numbers following each item show whether metre was duple, triple, or quadruple.

Special appendix two (statistical music) follows on page vi.

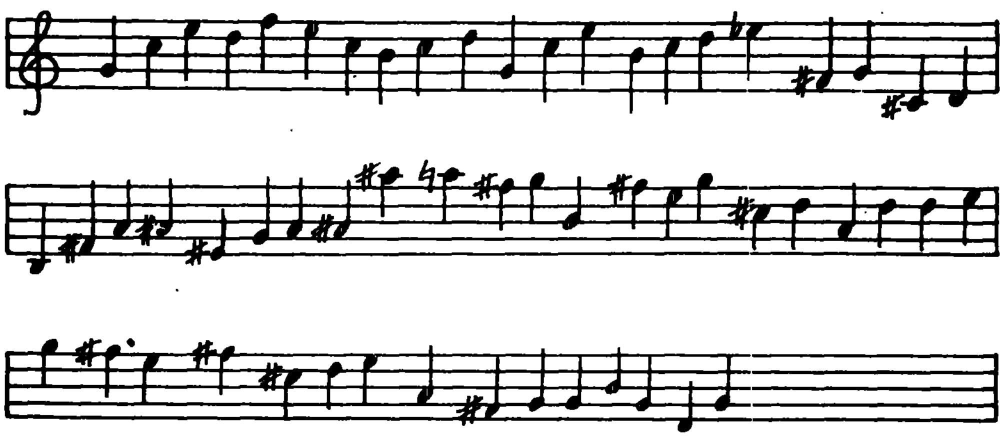
Practice Items.

Test Items.



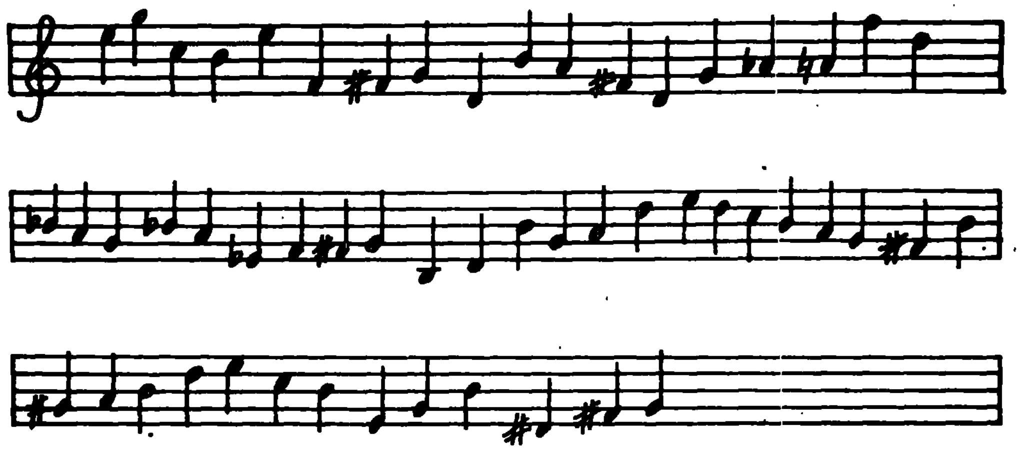
Examples of statistically derived approximations to music.

SEVEN NOTES RESTRAINT

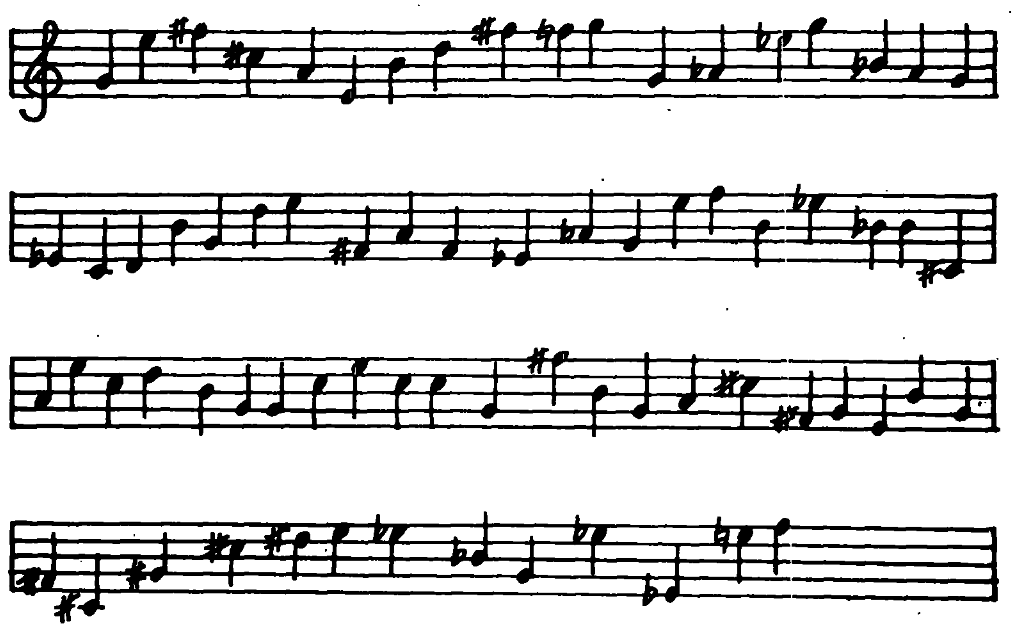


FIVE NOTES RESTRAINT

THREE NOTES RESTRAINT



ONE NOTE RESTRAINT



FINAL VERSION OF THE TEST BLANK.

NEW TESTS OF MUSICAL APTITUDE

J.B. DAVIES. DEPT OF PSYCHOLOGY, DURHAM UNIV.

PLEASE COMPLETE THE FOLLOWING:-

NAME ____________________

AGE _ yrs __ mths ______

SEX __________

TODAY'S DATE __________

SCHOOL ____________________

CLASS __________

TOWN ____________________

Tests of Musical Aptitude

Part 1

Practice

1	N	Y ✓	1
2	N ✓	Y	
3	N	Y ✓	2

Test Items

1	N	Y	
2	N	Y	
3	N	Y	
4	N	Y	
5	N	Y	
6	N	Y	
7	N	Y	
8	N	Y	
9	N	Y	
10	N	Y	
11	N	Y	
12	N	Y	
13	N	Y	
14	N	Y	
15	N	Y	

Part 2

Practice

1	1 ✓	2	N
2	1	2	N ✓
3	1	2 ✓	N

Test Items

1	1	2	N
2	1	2	N
3	1	2	N
4	1	2	N
5	1	2	N
6	1	2	N
7	1	2	N
8	1	2	N
9	1	2	N
10	1	2	N
11	1	2	N
12	1	2	N
13	1	2	N
14	1	2	N
15	1	2	N

TESTS OF MUSICAL APTITUDE

PART 3

PRACTICE

1	1	2 ✓	S
2	1 ✓	2	S
3	1	2	S ✓

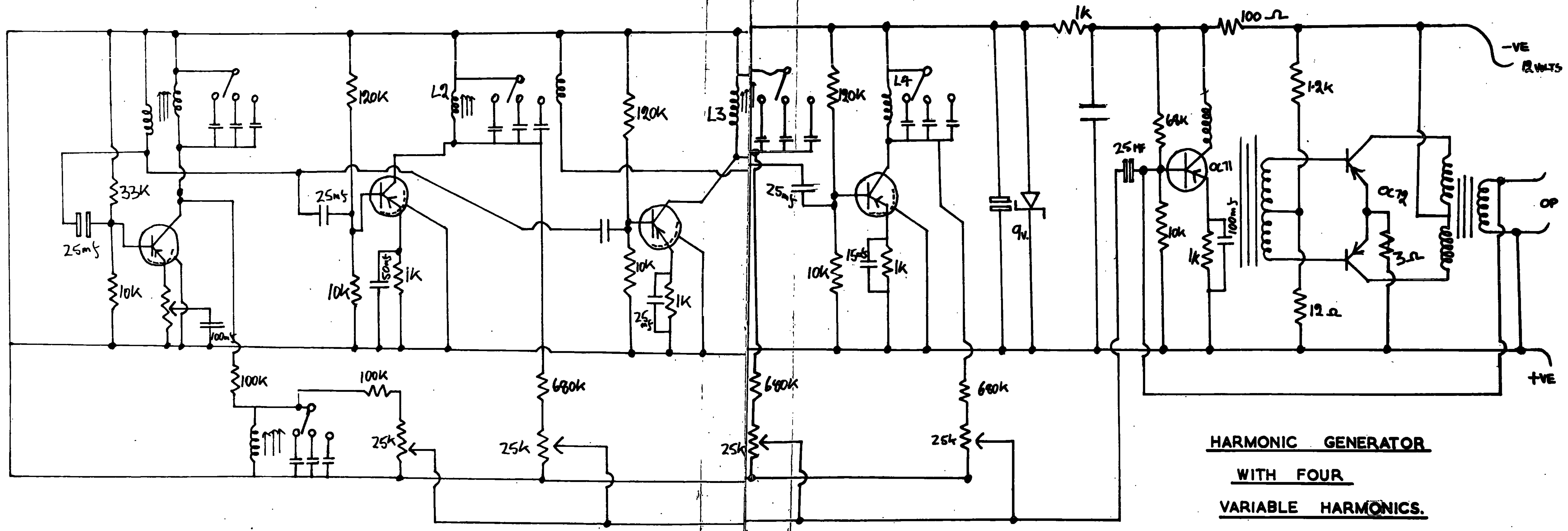
1	1	2	S
2	1	2	S
3	1	2	S
4	1	2	S
5	1	2	S
6	1	2	S
7	1	2	S
8	1	2	S
9	1	2	S
10	1	2	S
11	1	2	S
12	1	2	S
13	1	2	S
14	1	2	S
15	1	2	S

PART 4

PRACTICE

1	2	Y ✓	N
2	4	Y	N ✓
3	3	Y ✓	N

1		Y	N
2		Y	N
3		Y	N
4		Y	N
5		Y	N
6		Y	N
7		Y	N
8		Y	N
9		Y	N
10		Y	N
11		Y	N
12		Y	N
13		Y	N
14		Y	N
15		Y	N



HARMONIC GENERATOR

WITH FOUR

VARIABLE HARMONICS.

fig 4