This report defines musical aptitude in order to construct a test and obtain data pertinent to the perfecting of the definition and helpful in the practical assessment of student aptitude. Criteria for assessing potential music students often reflect achievement rather than aptitude; objective tests are often too atomistic and narrow in scope or are too subjective and dependent on culture and training. By defining musical aptitude as the ability to conceive the structure of acoustic material, and by placing that definition within a model representing the areas of important variable correlation with musical comprehension--areas of sensory discrimination, musical aptitude and some correlated factors, and personality--a basis for a culture-free test of aptitude was derived. The test was administered to 30 music institute students, ages 10-18, and consisted of recognition of a second series of sounds as a segment of a first series of sounds. Analysis of the results, which correlated .60 with teacher's ratings, have motivated further study which will consist of a lengthened test, the use of the test in selecting institute students, and the correlating of musical ability with general intelligence and spatial ability. (JH)
THE ABILITY TO STRUCTURE
ACOUSTIC MATERIAL AS A MEASURE
OF MUSICAL APPTITUDE
I. Background Theory and Pilot Studies
Kai Karma

THE ABILITY TO STRUCTURE ACOUSTIC MATERIAL
AS A MEASURE OF MUSICAL APTITUDE

I. Background Theory and Pilot Studies

Institute of Education
University of Helsinki
1973
Preface

The principals of the music institutes of Espoo and Kirkkonummi - Erkki Pohjola and Risto Saari - have kindly given possibilities of testing the ideas in practice.

Professor Emeritus Matti Koskenniemi, Dr. Juhani Jussila and Mr. Tor Kronlund have read the manuscript and given valuable comments of it.

Mr. Kari Äikäs has helped in making the test tapes.

I wish to thank all the forementioned; in all, I am thankful to all who have been tolerant enough to listen at my theorizing.

Helsinki, November 1973

Kai Karma
Contents

1. Introduction ........................................ 1
2. Music tests - discussion and critique .......... 2
3. Musical aptitude ................................. 6
   3.1. An overview ............................... 6
   3.2. Outlines of a theory .................... 8
4. Pilot studies .................................... 13
   4.1. Requirements set for the test .......... 13
   4.2. Testing of preliminary versions .... 14

References ........................................... 19
1. Introduction

The aims of this study are both theoretical and practical. Attempts to clarify the concepts concerning musical aptitude and their relations serve as a basis for a test which on the other hand gives possibility to validate empirically the theoretical considerations. This interplay between theory and practice seems to be too often forgotten in developing measures of musical aptitude. Theories are developed without sufficient connections with empirical facts and, especially, tests are composed with little knowledge about what is to be measured. This situation has been remarked by many writers, for example Bentley (1966, 14, 19) and Lehman (1968, 8).

In an ideal case anybody wanting musical education would have an opportunity of getting it. Unfortunately this is usually very far from the actual situation. Music classes in schools and music institutions have to select their pupils. Criteria in selecting pupils are of primary importance. When elementary musical training is considered an applicant should be judged according to his capacity to learn music and not according to the level he already has reached. Here tests with good predictive validity are sorely needed.

As mentioned above, tests often lack sufficient theoretical frame of reference. This causes "shotgun empiricism" which should be avoided. Thus the outlines of a theory on the nature of musical aptitude are sketched in this study. The model has been kept as general as possible to avoid the narrow view from one musical culture only.
The words "aptitude" and "ability" are used in this study in the meanings Lehman uses them (1968, 8). Thus "aptitude" is the capacity of an individual which affects his possibilities to gain "ability". Aptitude is the primary target of this study.

2. Music tests - discussion and critique

In this chapter single tests are not of primary importance but rather types of tests. Interest is centered on aptitude tests, achievement tests are not discussed here. It seems that testing artistic abilities - including musical ability - has not been developed as far as intelligence testing. There is not, however, any clear and decisive reason why it should be so. Very probably there is much that can be done in the area.

There are many ways of grouping existing tests. Here the most convenient point of departure seems to be the importance that is given to sensory capacities. In some tests sensory discrimination is almost the only object of interest, in others atomistic, sensory approach is avoided and more musical material is used.

The reasoning behind the tests of the first type seems to be mainly based on two assumptions:

1) Sensory discrimination is an essential part of musical aptitude.
2) Sensory capacities can not much be affected by training, i.e., they predict a person's future ability relatively well.
Empirical evidence doesn't give a clear answer to the question about the validity of these assumptions. There is, however, so much evidence against them that they must be strongly suspected. For example, Lundin (1967, 241) mentions studies where intelligence tests have predicted success in music studies better than Seashore's Measures of Musical Talents. This test has, instead, proved helpful in selecting people to certain works requiring auditory discrimination such as those of sonar operator and radiotelegrapher (Anastasi 1961, 410). Practical evidence also shows that sensory discrimination can be trained in a considerable degree. Tests of this type are often criticized because of their atomistic approach. Isolated pairs of sounds are used as basic material and the importance of the relations between several tones has been forgotten. Ability to differentiate tones which have small differences in pitch, length, intensity and timbre is measured relatively reliably but only one in some degree necessary but not at all sufficient condition for success in music has been reached.

These shortcomings have been tried to avoid in the tests of the second type. Small phrases of music often played with real instruments are here typical. These tests often include subtests where aesthetic preference is measured. There are some qualities in these tests which may be sources of serious faults:

1) Previous training in music tends to affect the results too much. This makes a test more or less like an achievement test instead of an aptitude test. Abilities which are directly trained in musical schooling are often measured. One such ability is the sense of tonality. It is good to remember that scales and cadences are usually an integrated part of music training. A trained person very probably has an unfair
advantage of this when the sense of tonality is measured. Tests where chords are analyzed - either by telling how many tones they consist of or which the tones are - are very probably affected by training, too. It seems that it is much more difficult to an untrained person to specify the pitches or number of tones in a chord even if he can easily differentiate the chords from each other according to their general quality.

2) A person's cultural background and interests tend to affect the results too much. The tests have properties which make them easier for persons with experience and training in traditional, western, tonal music. For example the sense of tonality is sometimes used in tests as if it were some kind of a biological property belonging to any musical person in spite of his experience and the culture he belongs to. Apparently, a good sense of tonality is an indicator of musical ability but the lack of it can not always be thought to indicate a lack of musical aptitude. The reason may as well be lack of training and experience. It may be thought that there is a more abstract property or aptitude which develops into the sense of tonality when certain condition are met. This property could be for example the ability to structure groups of sounds, to conceive their relations to a wider frame of reference. If this aptitude could be measured the results would probably be more culture-free and more interesting as theoretical findings than the results of existing tests usually are. To take an analogy from a different area: mechanical or technical ability can hardly be thought to be a primary, in a sense biological, mental ability. It is easier to think that some kind of spatial reasoning is the basis which can develop into mechanical ability in certain conditions.
3) The tests often include very subjective subtests. Aesthetic preference or the suitable completion to a (usually tonal) melody are of this type. It is sometimes claimed that objectivity and validity are tested by making trained musicians rate the items and including in the test only items with a sufficient degree of agreement. There is, however, a source of error in this method. Just because the raters are trained musicians they may be unable to know how an untrained - although musical - person feels about the items. It is very likely that this method tends to favour the traditional western type of concept of music on the cost of others.

4) Some tests seem to correlate too highly with intelligence. This may be caused for example by difficult instructions.

One shortcoming common to almost all the tests is the focusing into their reliability on the cost of validity. This is probably partly caused by the fact that reliability is a relatively simple psychometric property which can be directly computed from the original data while the determination of validity requires further investigations. Attempts to reach a high reliability should not be made on the cost of validity; a test with a high reliability may be completely useless. It is quite realistic to suppose that some tests supposed to measure musical aptitude actually measure to a great extent "knowledge about the western tradition" or "attitudes towards certain ways of playing".

Trying to get a high reliability often causes other faults, too. Because the reliability of a test depends among other things on the length of the test, tests are often made too long and boring. It must be remembered that the length of a test has a much smaller effect on validity than on reliability (e.g. Cronbach 1966, 131). A high reliability may
also give a misleading picture to a person who does not exactly know the meanings of the terms.

As a conclusion it may be said that the abstract, objective tests are often mechanical, atomistic and narrow in scope while attempts to bring the test closer to music often cause subjectivity and dependence of culture and training. An ideal test might be in between: abstract enough to be objective and as culture-free as possible but so close to music that interest and sufficient wideness of scope are maintained.

3. Musical aptitude

3.1. An overview

It is difficult to give a good general definition of musical aptitude. More commonly, examples of necessary conditions for musical aptitude are named instead of a general definition. This situation is not good when testing of musical aptitude is considered. It causes terminological uncertainty and disagreement upon aims and methods in testing. Some properties often mentioned as important parts of musicality are discussed here to form a frame of reference to later considerations.

Discrimination of pitch is one of the most usual criterions of musicality, especially in everyday, non-scientific talk. Researchers do not agree upon its importance. Seashore, e.g., treats it as an important indicator of musical aptitude while Révész and Vidor, on the other hand, do not give it much value as such (Roiha 1965, 56 - 57). Probably the disagreement is partly due to difference in terminology: the latter may think pitch differentiation is important but do not include it into musical ability but some more mechanical,
sensory capacity. As mentioned before, practical experience shows that this ability can be much affected by training. This diminishes its efficiency as a predictor. It must also be kept in mind that in different tasks in the field of music the importance of pitch discrimination ability varies.

**Sense of rhythm** is usually considered important to good performance and understanding of music. According to Roiha a good sense of rhythm is a condition to good musicality but not a sure indicator of it (Roiha 1965, 71).

**Sense of tonality.** According to Révész and Vidor this aptitude is very important (Roiha 1965, 75). The meaning of the term is not quite exact: it may be bound to tonal music and thus be considered as a sense of the major and minor tonalities or, which seems more reasonable, be treated more freely as a general sense of the relations and meanings of tones according to their pitches. Franklin (1969, 19) suggests that the term should be used in its narrow, clear sense and gives the term focusness to mean the wider concept.

**Musical memory.** Most researchers consider musical memory as a necessary condition of good musicality. The important point here is that it is not known if musical memory is a primary aptitude as such or if it is a consequence of understanding the material that is heard. It is much easier to remember well understood material than one only superficially heard.

**Emotional-aesthetic musicality** is often considered as one of the most important indicators of musicality or even treated as synonymous with it. According to Roiha “differences in emotional capacity and the direction of attitudes towards music are often more important than small differ-
ences in basic musical talents" (Roiha 1965, 95).

**Gestalt-forming ability.** The ability to divide the material into meaningful subgroups, to conceive the different importance of the tones as well as the classic example of gestalt psychology: transposition are examples of this category. It is useful to note here that a gestalt may be either formed in accordance of objective differences in the material or subjectively perceived in a group of objectively similar stimuli.

3.2. Outlines of a theory

In defining "musical aptitude" a definition of "music" would be of help. The definition should be as general as possible to avoid centering the interest into some narrow area only. Even if Varèse's definition "music is organized sound" is no real definition - there is organized sound which is not music and music which is not meant to be organized - it gives a frame of reference to thoughts. In a general sense there is some kind of organization in any music, in fact it exists at once when two or more sounds are in connection with each other. It can be said that the structure formed by the sounds is the music itself, i.e., it is not essential what sounds are used but what are their relations to each other. Transposing is a good example of this: a transposed piece of music may not include any of the original sounds but is still considered the same piece. It seems reasonable to think that conceiving this structure is an essential condition to an adequate apprehending of music. This is in accordance with the structuralist theory which assumes that the world of representations can be divided into small pieces, elements of structure, which are
put together in a definite way. The assembly of the elements is itself the structure ... Any organism, including a machine, is a structure; to understand a machine is to perceive that structure" (Moles 1966, 20, 33).

In this phase a preliminary definition of musical aptitude can be given: musical aptitude is the ability to conceive the structure of acoustic material. The purpose of the rest of this chapter is making this definition more accurate and bringing it into an empirically testable level.

The ability to conceive the structure of a group of sounds is of course not a dichotomous variable. The structure may be apprehended also partially. Thus it is convenient to operationalize the ability as follows: the ability to conceive the structure of a group of sounds is equal to the amount of information about the relations of the sounds a person gets when hearing them. The statement "amount of information" gives us a possibility to quantify the ability.

Let us consider the situation with the following simple example:

\[ \text{\includegraphics[width=0.5\textwidth]{example.png}} \]

The amount of information may vary within very wide limits. If all a person can say about this group of sounds is that it consists of three sounds of different kind, he may be relatively safely judged as unmusical. There is a little more information in the perception "the group consists of three sounds varying in pitch and length". When still more information is added the following phase is reached: "there
are three sounds, the first being the lowest and the last the highest. The first two are of same length, the third is longer". It seems reasonable to expect this from a musical but not trained person. For more accurate comprehending, schooling begins to be necessary. An example of this could be for example the following: "There are three sounds, the second being at a distance of minor third from the first and the last at a distance of major third from the second. The last sound is twice as long as the two first sounds".

It is not necessarily meant here that the comprehension should be conscious; that it could be verbally explained. The forementioned examples just describe the amount of information the person operates with. The actual comprehension may as well be inconscious and intuitive.

Musical aptitude is considered in this work a pure aptitude variable, personality traits are not included in the concept. This does not mean that personality would be considered a less important factor but that it can be separated from musical ability. Keeping them apart is clearer from the point of view of research. If, according to Cronbach (1966, 29), aptitude variables are defined as the maximal performance of an individual and personality variables as his typical behaviour, the difference is relatively clear. It could be roughly said that the way a person can comprehend or perform music depends on his musical aptitude and how he typically does or wants to do it is mainly due to his personality.

Sensory capacities are also considered important factors of musical perception and performance but are not included in the concept of musical aptitude.
Defining musical aptitude in the way described above is beneficial in the sense that it is economic and includes many usually mentioned criterions of musical aptitude as special cases. For example the senses of rhythm and tonality are ways of conceiving a structure, comprehending the relations of sounds. Structuring ability may thus be thought of as a hypothetical construction, a background variable which is used to explain empirical facts. For example "intelligence" is a similar construction which is used as a general explanation for the differences in problem-solving efficiency between people. A test measuring directly this background variable, in this case structuring ability, would be an ideal measure of musical aptitude. If this could be done, measuring parallel phenomenons greatly affected by training and experience, such as the sense of tonality, would not be needed. This short cut is, of course, of help only if there is one general ability which explains a great deal of the variance in musicality between people. If musical aptitude consists of several relatively independent parts, the meaning of this model is greatly diminished.

It is interesting here to note that there is a close connection between visual spatial factor (S) and musical aptitude such as it is defined before. Both are abilities to structure or organize one's perceptions, although the channel of information is different. It is here supposed that the similarity is not only superficial, but has a real basis in similar neural functions. Thus these abilities should correlate positively.

The considerations presented here may be clarified with the following figure:
Figure 1. A model representing the relations between the most important variables correlated with musical comprehension.

Hypothetical constructions

<table>
<thead>
<tr>
<th>Discrimination of pitch</th>
<th>Seashore-type measures, ability to produce exact intervals etc.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discrimination of timbre</td>
<td></td>
</tr>
<tr>
<td>Discrimination of intensity</td>
<td></td>
</tr>
<tr>
<td>Discrimination of time</td>
<td></td>
</tr>
</tbody>
</table>

Operational level

<table>
<thead>
<tr>
<th>Ability to structure sounds = musical aptitude</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intelligence</td>
</tr>
<tr>
<td>----------------</td>
</tr>
<tr>
<td>Sense of tonality, sense of rhythm, musical memory etc.</td>
</tr>
<tr>
<td>Musical taste, attitudes towards music, etc.</td>
</tr>
</tbody>
</table>

Area A: Sensory discrimination

Area B: Musical aptitude and some correlated factors

Area C: Personality

- Temperament
- Sensitivity
- Introversion
- Etc.
The model implies several assumptions:

1) The sense of tonality, the sense of rhythm, musical memory etc. are based on one general ability to structure one's perceptions of sounds (=musical aptitude).

2) Sensory capacities are separable from musical aptitude.

3) Sensory capacities are separable from each other.

4) Personality traits are separable from musical aptitude.

5) Musical aptitude correlates more positively with spatial ability than general intelligence.

The purpose of this study is to investigate area B by means of a test developed to measure the ability to conceive the structures of groups of sounds.

4. Pilot studies

4.1. Requirements set for the test

When test construction began, several requirements were set to guide the construction. Most of these were direct consequences from the critique described before. The main requirements were the following:

1) The test should be as culture-free as possible. This
is tried to achieve by making the test so abstract that it does not include actual music or material from some musical subcultures on the cost of others.

2) The test should be as free as possible from the effects of music training. The test should not include material directly practiced in music schools or lessons.

3) The test should be objective. There should be only one right answer to every item. The answer should be such that it can not be disagreed upon.

4) The test must not be too long and boring.

5) The test must not be affected by the differences in sensory discrimination ability within subjects.

6) The test must be suitable for as young subjects as possible.

4.2. Testing of preliminary versions

It was said before that the ability to conceive the structure of acoustic material was operationalized by substituting it with "the amount of information etc.". This definition is not, however, suitable for practical test construction as such. It does not give the actual practical operations which are needed to measure the ability. It must be further operationalized when making a test is considered. This was tried with three different kinds of items: first, transposing, second, choosing a suitable completion to a figure of sounds developing in a certain way and third, dividing a series of sounds into subgroups. Part of the material was produced with
the "dimi"-syntethiser of the Institute of Musicology of the University of Helsinki, part with an electric organ. The subjects were 10-year old boys and girls from an ordinary elementary school class.

The experiences from these preliminary versions led to the following conclusions:

1) The motivation of the subjects is very important. Items produced with the synthesiser were considered very boring because of the monotonous tone of the machine. If synthesiser is used its tone must be altered from item to item by filtering, adding vibrato etc.

2) The time interval in the items and between them were too long. The tones may be presented much faster than the maker of the test feels. The same phenomenon has been noted by Petzold (1966, 20).

3) The way in which the answers are given should be the same through the test. Changing, e.g., from multiple choice type to dichotomous items tends to confuse some subjects. The danger of measuring intelligence is thus very near.

4) The items must be very carefully studied to avoid measuring reasoning-type abilities. Problems where a completion to a developing series of sounds was to be found, clearly correlated too much with reasoning ability operationalized with Raven's Progressive Matrices and sentence completion tests.

The test where series of sounds were divided into subgroups seemed to be the best one. Thus, this test was chosen to form a basis for later attempts and is described here more in detail.
Every item consists of two series of sounds. The first one is longer and is constructed so that it can be divided into three similar parts. The second, "answer" series follows the first one after three seconds.

Example:

<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The subjects have to decide if one third of the first series is similar to the "answer" part, i.e., if the shorter group would form the longer one when repeated three times. The subjects answer by making a ring around "yes" or "no" in their papers. This type of answering was chosen instead of multiple choice items for two reasons: first, it is clearer to young subjects and second, it makes it difficult to signal the answers to other subjects. If multiple choice type is used the right answer can be often found just by looking when others mark their papers.

The test consists of 27 items and lasts about fifteen minutes including the instruction. Only one parameter is varying in any item the others being constant. Thus, if e.g. intensity varies, pitch, timbre and length are held constant through the item. This makes the test a little more "unmusical" but is of help when the results are interpreted. Items where several factors vary at the same time may be used in later versions when basic experience of tests of this type has been attained.

The timbre varies between the items to maintain the subjects' interest.
When an item begins the subjects do not know how many units it consists of. If the expected amount of units is not appropriate to the rest of the series the structure first conceived has to be quickly reshaped. Thus it may be said that the test measures, in addition to structuring ability as such, also the flexibility of structure forming, the ability to change conceived structures. According to the practical experience of the author this may be a good indicator of musical aptitude.

When the test had been given to a group of 30 music institute students from 10 to 18 years of age an item analysis was made by correlating the items with the whole test. Some items had an almost zero correlation and were removed. The rest of the test had a Kuder-Richardson-reliability of .68. Teachers' ratings about the pupils' musicality correlated .60 with the test results. The teachers were asked to rate their pupils according to musical aptitude, not sensory capacities or the present skill attained by training. In this preliminary trial there was no remarkable tendency of the items to group according to the parameter varying in them. In all, the results were in accordance with the considerations made about musical aptitude and thus motivated further studies.

In the next phase of the study the following operations will be made:

1) The test will be lengthened to consist of about 40 items.

2) The test will be used in connection with selecting applicants to a couple of music institutions to compare it with the selecting methods used and to get a possibility of investigating its predictive validity.
3) The test will be used with tests of general intelligence and spatial ability to validate the hypotheses made about their possible relations with musical ability.
References:


Previous Issues in this Series

No. 1  The Effectiveness of Punishments and the School Climate by OLLI SIPINEN. May, 1957. 10 pp. Out of print.


No. 3  On Deceit Practiced in Finnish Grammar Schools at Written Tasks by OALVI PAIVANSAKO. Nov., 1957. 5 pp. Out of print.

No. 4  Comparison between Reading Comprehension and Reading Aloud by KAIJA HÄLINEN. March, 1958. 9 pp. Out of print.


No. 6  The Awarding of School Marks by TOIVO VAHERVUO. Sept., 1958. 5 pp. Out of print.


No. 8  Motives for the Choice of the Teaching Profession Correlated with Success in the Entrance Examination and in Student Teaching by MATTI KOSKENNIEMI. Feb., 1960. 11 pp. Out of print.


No. 16  Some Problems in Combining the Individual Tutors' Marks into a Final Mark in Teaching Ability by JOHANNES ALIKOSKI. March, 1966. 25 pp. Out of print.


No. 20  On the Anxiety Associated with School Attendance and the Grammar School Entrance Examination by EEVA PATJAS. May, 1968. 16 pp.


No. 30 Investigations into the Instructional Process. V. Experiences with the Bellack Classification System by KAI KARMA. Jan., 1972. 16 pp.


