

RESPONSES TO MUSIC

Thesis submitted for the degree of Ph.D.

by Kate Castell, 1984.

FOR PAUL

## CONTENTS

### PART 1: GENERAL INTRODUCTION

1.1: Historical Perspective.....	1
1.2: Methodological Issues.....	7

### PART 2: DEVELOPMENTAL RESEARCH IN THE PSYCHOLOGY OF MUSIC

2.1: Literature Review.....	17
2.2: The Present Research.....	42
2.3: Experiment 1.....	49
Method.....	49
Analysis of Responses and Results.....	53
Discussion.....	60
2.4: Experiment 2.....	79
Method.....	81
Analysis of Responses and Results.....	83
Discussion.....	86
2.5: Conclusions (Part 2).....	101

### PART 3: AESTHETIC REACTIONS TO MUSIC: A COGNITIVE/NATURALISTIC APPROACH

3.1: Literature Review.....	105
3.2: The Present Research.....	115
3.3: Experiment 3.....	122
Method.....	122
Analysis of Responses and Results.....	130
Discussion.....	134

3.4: Experiment 4.....	152
Method.....	152
Analysis of Responses and Results.....	156
Discussion.....	160
3.5: Experiment 5.....	181
Method.....	181
Analysis of Responses and Results.....	185
Discussion.....	186
3.6: Conclusions (Part 3).....	201

#### PART 4: LABELLING OF MUSICAL STYLES: A COGNITIVE/NATURALISTIC APPROACH

4.1: The Role of Categorisation in Music Perception.....	207
4.2: The Present Research.....	219
4.3: Experiment 6.....	223
Method.....	223
Analysis of Responses and Results.....	226
Discussion.....	227
4.4: Conclusions (Part 4).....	241

#### PART 5: THE EFFECTS OF EXTRA-MUSICAL INFORMATION ON RESPONSES TO MUSIC

5.1: Some Previous Research.....	244
5.2: The Present Research.....	251
5.3: Experiment 7.....	254
Method.....	254
Analysis of Responses and Results.....	258
Discussion.....	259



PART 6: GENERAL CONCLUSIONS

6.1: Summary of Findings..... 270

6.2: Implications and Suggestions for Further Research:

    The Concept of Musical Style..... 275

APPENDICES..... 285

REFERENCES..... 305

## ACKNOWLEDGEMENTS

I should like to thank all those who have helped me with this thesis, especially David Hargreaves for his unobtrusive supervising: his advice was indispensable and always willingly given. I am very grateful to the SSRC who financed the research, and to Diane Williams and Nic Trower who typed it up. Thanks also to Paul Riddick who helped to collect the data for Experiment 6, to my mother for help with proof-reading, and to Paul Windridge for his draughtsmanship and moral support.

Most of the work described here could not have been carried out without the cooperation of the schools mentioned below, and I am indebted to all the staff and pupils who were involved:

Beaumont Leys Secondary School, Leicester.

Great Glen Primary School, Leicestershire.

Hazel Street Primary School, Leicester.

Oadby Beauchamp School, Leicester.

Wigston All Saints' C.E. Primary School, Leicester.

#### NOTE

In Experiment 6, the word "Muzak" is used as a label for a musical style. After the study had been completed, it was pointed out to the author that this word is in fact a trademark. Although in this respect this label was not quite equivalent to all the others used, it was felt that it nevertheless provided as good a description of a certain style of music as did the other labels.

## PART 1: GENERAL INTRODUCTION

### 1.1: Historical Perspective

Human beings' responses to music are infinitely varied, not only in their content but also with respect to the medium in which they occur. In any one person, different pieces of music may produce an urge to dance, a feeling of intense melancholy, boredom, incomprehension, the feeling of having had a sudden and inexpressible insight, a vivid evocation of some past experience, or simply a feeling of comfortable familiarity. All these may or may not be observable reactions, depending upon whether or not the person decides to dance, weep, drum their fingers on a table-top, look puzzled, or talk about how they feel, respectively: and these represent but a few of the possible responses to music that person could make. To complicate the issue further, the same piece of music might conceivably give rise to very different responses in the same person under different circumstances, whether social, environmental, or purely affective; and responses are furthermore liable to change dramatically with age and experience.

All these variables: the nature of the response, the means of its expression, the extramusical variables which might have influenced it, and also the internal

characteristics of the music that gave rise to it: are potential subject matter for investigators in the field of music psychology. Until relatively recently the area remained largely unexplored from a psychologist's point of view, but from about 1960 onwards there has been a proliferation of studies into all aspects of the relationship between the perceiver of music and the music that is perceived.

The first section of this thesis will mention some of the research which has led up to the present state of affairs, via a description of some aspects of musical ability testing. This approach is adopted because (a) much of the earliest research in music psychology tended to be centred around the development of tests of musical aptitude and ability; and (b) issues brought to light through the problems encountered when designing ability tests echo, in essence, issues that emerge as important in the field of music psychology as a whole.

Some of the earliest sustained research in music psychology was carried out by Seashore (1938). He was particularly interested in musical ability and aptitude, and constructed a battery of tests (Seashore 1919, 1939, 1960) which reflected in their subdivisions his belief that musical ability is not a unitary or general characteristic but is made up of various components such as pitch discrimination, timbre discrimination, sense of time, and sense of consonance. He felt that: "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'." (1938, p.2). The "musical mind" he

described as a mind:

...capable of sensing sounds, of imagining these sounds in reproductive and creative imagination, of being aroused by them emotionally, of being capable of sustained thinking in terms of these experiences, and ordinarily, though not necessarily, of giving some form of expression of them in musical performance or creative music. (1938, p.1).

To take the analogy further, Seashore believed that the roots of this tree are nourished by hereditary rather than environmental factors: in other words, that the building blocks of musicality are inborn and can be detected and measured by using the right tools. The hereditarian aspect of Seashore's work is important because it was the basis of his justification for the construction of aptitude tests: if a child was found to have no demonstrable musical aptitude, there was then little reason to spend money and effort in providing that child with a musical environment, whereas a child who performed well on aptitude tests would benefit from such an environment.

Two important issues are brought to light by Seashore's studies of the "musical mind". Firstly, he drew a distinction between the "physics" of musical events and the nature of the mental experiences corresponding to those events. For instance, he discussed the differences between frequency (which may be accurately measured by physical methods) and



pitch (which is the subjective experience of frequency), between intensity and loudness, between duration and the experience of time, and between wave form and timbre. In each of these pairs of instances, the first-mentioned is a physical or acoustical variable which may be objectively measured, and the second is the corresponding subjective experience. Thus Seashore was well aware of the notion that the subjective experience of music is not necessarily synonymous with objective measures of the sound itself.

The second issue applies to musical aptitude testing, and indeed to aptitude testing in general. The Seashore battery of tests does not use complex tones or actual musical instruments, but pure tones and other stimuli which are not biased in favour of musically experienced subjects: such a bias would immediately defeat the object of measuring musical aptitude rather than degree of learned musical ability. As Seashore wrote, "In a test program of this kind, we should stay as far as possible away from anything that involves musical training and experience, if we wish to predict success on the basis of talent." (1938, p. 310). Davies (1971) pointed out that Seashore is a notable exception in this respect, because most test authors have used material which is more explicitly musical and which in some cases consists of intact musical excerpts and structures (e.g. Wing 1948, Hevner & Landsbury 1935, Lowery 1926). Both Seashore and Davies are aware of the biases that can arise from using musical stimuli which are drawn exclusively from the Western classical repertoire, and as will become clear later in the present

thesis, this problem is not restricted to aptitude testing.

In 1971, Davies devised a series of tests which use nonmusical or quasi-musical stimuli rather than formal musical material. After administering the tests to over 2,000 school children he concluded that a battery of such tests was at least feasible. Because of the nonmusical or quasi-musical nature of the test items, the extent to which responses to tests like these are accurate predictors of musical responses and activities in later life has to be carefully checked; but as Davies pointed out, "Any situation can give a guide to the degree to which talent is present, if we can show that it meets certain criteria of validity and reliability." (p. 557).

There are various musical aptitude tests which might be described as "atomistic" in that they examine only specific facets of musical ability, as do the separate sections of Seashore's (1919, 1939, 1960) battery of tests. Examples of these are the Watkins-Farnum performance scale (1954), the Gordon Iowa Tests of Musical Literacy (1970), and the Colwell Music Achievement Tests (1969-70). Recently, however, there has been a move towards a "global" approach both in terms of testing and in music psychology research as a whole. For instance, Bullock (1973) reviewed tests which aim to assess a variety of perceiver attributes which are more general, and hence less easy to define, than specific abilities like the discrimination of pitch or of loudness. Bullock mentioned a number of variables which might contribute to what he described as a global "musico-aesthetic attitude", among which are mood, taste, interest, preference, perception, and



appreciation.

So far the discussion has cited only musical aptitude tests, but early research in music psychology did not consist solely of stimulus-based studies whose aim was to produce an efficient aptitude test. One of the earliest attempts to study some of the more subjective aspects of responses to music experimentally was probably that by Heinlein (1928), who examined the affective characteristics of major and minor modes. There was also research into the effects of music on the morale of workers (Halpin, 1943) and some investigation into factors affecting musical preference. Examples of studies in the latter area are those by Cattell and Saunders (1954), Fay and Middleton (1941), and Wiebe (1940), in which the researchers examined the influences on preference of personality, musical talent and radio "plugging" respectively. Most of these early studies employed natural music as their stimulus material, rather than using specially designed tone sequences in an effort to gain greater control over all the variables which might contribute to a particular aesthetic response. The latter trend of using non-natural music did not have a great impact, other than in the area of formal aptitude testing, until the 1960's, when it represented an important methodological development. The issue will be discussed in the next section of this thesis.

## 1.2: Methodological Issues

"Music is essentially useless; as life is: but both lend utility to their conditions." (George Santayana, 1920).

It was pointed out in the previous section that reactions to music, and indeed to any other art, are difficult to quantify. This is because the forms which they take may involve several psychological phenomena, either singly or interactively, which are themselves open to further investigation. Examples of such phenomena are emotions, attitudes, and implicit or explicit value judgements. These may be mediated by language, physiological changes, or other observable responses, but even if none of these occur it does not follow that there has been no reaction at all. Thus there may be a large number of widely divergent psychological processes involved in reacting to aesthetic stimuli.

It is appropriate at this point to give some explanation of what is meant by an aesthetic stimulus in the context of the present thesis. The term is used in a very broad sense, and refers to almost any entity which can be perceived in such a way that the perceiver's response to it has an affective component. This definition obviously covers all the traditional art forms, but it is also extended beyond this point. It could be argued that one factor held in common by works of art is that they are not intrinsically "useful"

objects, in the everyday sense of the word. (For instance, a Picasso painting might be useful to light a fire, but this function would be performed just as efficiently if a blank canvas were used: it is the canvas and wooden frame which are inflammable rather than the subject matter of the painting itself.) If the definition of an aesthetic stimulus included only objects which have no practical use, many things would be excluded which may be both useful and beautiful, such as a cathedral, a field of ripe wheat, a porcelain tea service, and even the human face. Even the most mundanely practical things can be considered beautiful, often as a direct result of their fitness to perform a specific function; as any motorbike owner who has just stripped down and rebuilt an engine will testify.

The concept of fitness implies a variety of parsimony: a well-designed engine may be seen as more aesthetically pleasing than one in which two or more components are used to do the job of one, thus increasing the risk of a break-down, just as a well-written computer program may be preferred to one in which several complicated subroutines have been added as afterthoughts, even though both are capable of carrying out the same task. Musical ideas, too, may be perceived as more or less economical or elegant in their expression, and to this extent criteria of intrinsic fitness or economy may be applied to pure art forms such as music: a sentiment better expressed by George Santayana in the quotation at the beginning of this section. Conversely, it should be apparent that in the present context, entities other than "pure" art works may be described as aesthetic stimuli.

It is clear that the experiences and responses of a perceiver of an aesthetic object are difficult to measure, and that the object of perception may also be extremely complex. With respect to the latter issue, any general theory of aesthetics would have to involve identification of, and predictions about, properties common to such diverse entities as a Mozart symphony, a pair of fashion shoes, an abstract painting, a flower arrangement, and a romantic novel, to name but a few examples. Two possible approaches to this problem have been adopted in the past. The first of these has been to select parameters along which any aesthetic object may vary, such as complexity or familiarity, and measure the effects on subjects' responses of systematic variations within these parameters. This approach was taken by Berlyne (1974), who coined the phrase "new experimental aesthetics" to describe it. It is stimulus-based, rather than perceiver-based, in the sense that it involves examination of the stimuli in question. For instance, a symphony or a pop song might be analysed to discover its constituent elements such as its tempo, rhythmic characteristics, melodic complexity, or timbre at any given point. A researcher might then proceed to account for the effects on the perceiver of the music in question, via an increased understanding of the nature of the elements of that music.

A second possible approach is to argue that musical phenomena can be understood only through scientific investigation of human processing mechanisms, and that psychological processes occurring within the perceiver are

paramount in determining the nature of an aesthetic response. This approach emphasises attributes of the perceiver, such as personality, age, intelligence, socioeconomic status, sex, and affective state.

As stated here, the two approaches appear to be diametrically opposed. However, this is an obvious oversimplification of the issue. Although the emphases in each case are different, the two approaches are complementary rather than contradictory: it would be nonsensical to try to understand how a piece of music is perceived without making any reference to the perceiver, just as it would be nonsensical to tabulate a perceiver's reactions without considering the piece of music which had given rise to them in the first place. The distinction between these two orientations, between studies in which the emphasis is upon stimulus attributes and those where perceiver attributes are examined, appears frequently when reviewing music psychology research, and there are also studies in which the two are combined to good effect. Before mentioning these, it is relevant here to introduce a second, related dichotomy which exists in the area, and which is again more hypothetical than real.

One of the difficulties involved in studying the stimulus characteristics of music is that music in general consists of highly complex auditory patterns which cannot meaningfully be scaled along simple physical stimulus dimensions. This makes it very awkward to manipulate as an independent variable. A similar problem exists for other art forms, and the solution



adopted by Berlyne (1974) was to use what he described as a "synthetic" approach. This consists of isolating particular variables or very simple elements of art works which might affect aesthetic appreciation, and designing stimulus patterns (for instance tone sequences, line drawings, rhythmic patterns etc.) in which these variables have been systematically manipulated. The effects of these manipulations on subjects' responses may then be readily observed. Examples of research in which this method has been employed are Crozier, (1974); McMullen, (1974); Konecni and Sargent-Pollock, (1976); Deutsch, (1982); and Pflederer, (1964, 1967).

This method has the drawback that the stimulus objects to which subjects are required to respond appear artificial and could not be described as "works of art", and as Conley (1981) argued, "The mainstream trend in studying musical perception has been to utilize the synthetic approach and to generate experimental results having no demonstrable ecological validity." (p.452). Thus although the approach enables an admirable degree of precision to be attained in the control of any independent variables, it is not possible to say to what extent any results obtained may be generalised to "real-life" situations. As Berlyne (1974) pointed out, the impact of a work of art depends on the combination and interaction of elements which may only be studied individually using a synthetic approach. The alternative course is to use genuine musical compositions, or portions thereof, and Berlyne described this as the "analytic" approach. This too has its attendant difficulties. Although it brings the experimental

situation closer to what happens in real life when people encounter works of art, there are very many possible interactions of a large number of elements within a given composition, regardless of any extrinsic variables such as its familiarity or the associations it evokes, that could affect the listener's responses. Consequently it is very difficult to even define, let alone measure, the effects of any one independent variable upon the aesthetic response produced. Nevertheless there have been many studies which have used real music as an aesthetic stimulus, among which are those by Conley (1981), Chapman and Williams (1976), and Getz (1966).

The distinction between the synthetic and analytic approaches has been commented upon by many workers; indeed, Berlyne (1974) pointed out that it was recognised by Fechner as early as 1876. However, not all of these researchers and theorists have used the same terminology. Meyer (1956), when discussing what he referred to as "atomism", was in fact outlining the problems associated with the synthetic approach when he said, "The attempt to explain and understand music as a succession of separable, discrete sounds and sound complexes is the error of atomism." (p.5). Deutsch (1982) stressed in the preface of The Psychology of Music that because of rapid developments both in the understanding of the nature of sound and in technology, which have enabled investigators to generate complex sound stimuli with versatility and precision, greater stimulus control has become possible. Broadly speaking, the emphasis throughout Deutsch's book is upon "...the relationship between the objective, physical

properties of sensory stimuli in our environment and the subjective, psychological responses evoked by them." (Rasch & Plomp, in Deutsch, 1982, p.1), and the range of subjects discussed may be more easily subsumed under the term "psychoacoustics" than "aesthetics" or "experimental aesthetics". In other words, the contributors to the book were concerned with making precise measurements and with carrying out careful manipulations of certain salient parameters which are constituents of acoustic stimuli, and the book represents the synthetic rather than the analytic end of the continuum. Sluckin, Hargreaves and Colman (1983) suggested that the synthetic and analytic approaches in experimental aesthetics might more appropriately be called "experimental" and "naturalistic" respectively, and this is the nomenclature which will be adopted throughout the remainder of this thesis. Some examples have already been given of research which has adopted one or the other approach, but studies in which the approach is exclusively experimental or exclusively naturalistic are actually relatively rare. The two approaches may more realistically be thought of as being at opposite ends of a continuum, with many studies falling somewhere in between the two poles because they make use of techniques characteristic of both approaches. Heyduk's (1975) study of the relationship between preference and musical complexity is worth mentioning at this point, because the stimulus material he used, although experimental in that it was specially constructed to represent different degrees of complexity, resembled naturalistic music much more than did the tone or



rhythm sequences of McMullen (1974) or McMullen and Arnold (1976). The latter are mentioned because they provide examples of unambiguously experimental stimuli. They are discussed in more detail in Part 3.1.

Thus it is possible to conceptualise any given piece of aesthetics research, and in particular research in the psychology of music, as existing at some hypothetical point on a two-dimensional model involving two continua; one from the experimental to the naturalistic approach, the other from an approach which emphasises stimulus attributes to one emphasising perceiver attributes. These continua are of course arbitrary and are not necessarily orthogonal: the experimental/naturalistic distinction is by definition more salient for research emphasising stimulus characteristics than it is for that emphasising perceiver characteristics. In addition, they are certainly not the only continua along which research may be categorised: for instance, the situation in which music is heard may be extremely important. However, they do provide a useful frame of reference from which to discuss relevant work because the advantages and disadvantages associated with each possible combination of approaches differ considerably. Because of this, when trying to obtain a realistic overview of the state of current findings, it is very important that research from all combinations of approaches should be represented, and an attempt has been made to do this in the present thesis.

There follows a summary of the contents of this thesis, described in terms of the model outlined in the preceding

paragraphs.

Part 2 examines some aspects of the influence of a perceiver attribute (age) on the extent to which excerpts of music sound stylistically similar. Because the musical stimuli used are naturalistic, it was not one of the aims of the experiment to investigate whether any specific stimulus characteristics were important in determining subjects' responses, but it was anticipated that an examination of the written justifications subjects gave would reveal some of the strategies used by them to decide whether or not two excerpts of music represented the same musical style.

Part 3 adopts an experimental approach, in the sense that non-naturalistic stimuli are involved. In Experiment 3, the emphasis is upon stimulus characteristics: specifically, the effects on similarity judgements are examined of changes in tempo or in rhythmic or melodic complexity. These variables were selected on the basis of subjects' responses in Experiments 1 and 2. In Experiments 4 and 5, some perceiver attributes (namely, subjective estimates of musical complexity, and the effects of an imagined mood) are also investigated, using the same type of experimental stimuli as in Experiment 3.

Part 4 further explores the concept of musical style, using naturalistic stimuli. On this occasion the issue at stake is the interaction of stimulus and perceiver characteristics. The musical stimuli were chosen on the basis of their stylistic ambiguity, and the aim of the experiment was to investigate the way in which subjects attempt to apply

stylistic labels for such excerpts. There was also some manipulation of the subject population, age being an independent variable.

The musical stimuli used in Experiment 7 were again naturalistic. However, some degree of control was attained over stimulus characteristics, through the selection of four excerpts, each of which was likely to be perceived as representative of both jazz and classical styles of music. The extent to which the perceived style of any music is actually a reflection of physical stimulus characteristics is arguable, and this issue was investigated by playing to two groups of subjects excerpts which were physically identical but which in one case were described as "modern progressive rock" and in the other as "contemporary classical". Age was a second independent variable.

Overall it can be said that no one methodological strategy will lead to a complete understanding of how people respond to music, if indeed such a goal is ever attainable. There are so many variables which must be taken into account that to have even a small degree of predictive success, a model of the mechanisms of music perception must incorporate findings from as many approaches as possible. The development of such a model is no easy task, and the present thesis makes no claims to attempt it. It is hoped however that the research described here will make some contribution towards the realisation of this aim.

## PART 2: DEVELOPMENTAL RESEARCH IN THE PSYCHOLOGY OF MUSIC

### 2.1: Literature Review

It is interesting that much of the research which investigates the development of children's responses to music has a distinctly educational and practical orientation, as illustrated by the existence of such journals as the "Journal of Research in Music Education" in the U.S.A., and "Psychology of Music" (the journal of the Society for Research in Psychology of Music and Music Education) in Britain. To take some examples, there are a large number of studies concerned with teaching methods (Hoffer, 1982; Madsen, 1982; Jorgensen, 1980), the place of music in the school curriculum (Plummeridge, 1980; Cleall, 1981), and the factors contributing to being a successful music teacher (Stuart, 1981; Kemp, 1982); and what all these studies have in common is that they were carried out to illuminate a specific practical problem. The resulting information is therefore rather "bitty": although many worthwhile findings have emerged, there is no cohesive theoretical background which can relate one finding to another.

In contrast, some other areas of developmental psychology appear (in retrospect at least) to involve a different approach because the contributing studies have been

theory-based in conception. The prime exemplar of this approach was Piaget, whose ideas have permeated the field from moral development to the growth of logical thinking and concept formation, but make very little reference to the development of aesthetic appreciation in general, including music appreciation.

There are of course many exceptions to this generalisation: for instance, several workers have investigated the extent to which Piaget's theories of conservation may be applicable to the way a child responds to music (Pfleiderer, 1964 and 1967; Larsen, 1973; Botvin, 1974; Serafine, 1979 and 1980; Nelson, 1980). In any case the distinction between the approach to the study of musical development and that to the study of development in other areas (e.g. science, mathematics, and language) may be a reflection simply of the fact that less work has been carried out in the former case than in the latter, so that research is still at a stage where information-gathering is very important and theorising on the scale of Piaget's would be inappropriate.

Although it amounts almost to a tautology to say that any developmental research might have implications for educational practice, the research to be discussed in this literature review will be that which is not primarily education-based because much of this is not strictly psychological, and is hence outside the scope of this thesis. The review is not comprehensive; instead specific studies are mentioned because they illustrate a particular avenue of investigation.



There is comparatively little research into the way very young children perceive and respond to music. A commonly held belief is that babies find any continuous, simple rhythmic pattern soothing, especially if it is in 6/8 time, because of the similarity of this sound to a heart-beat to which they have supposedly been imprinted during pregnancy. However, this has not been conclusively verified: although hearing does develop before birth it is not certain that it is possible for a baby in the womb to hear its mother's heart beating (Querleu, Renard and Crepin, 1981), although an unborn child does certainly react to loud noises outside the womb (Spelt, 1948). It has been observed (Stirnimann, 1940) that during the last few months of pregnancy a mother may notice that her unborn baby is particularly active when she (the mother) listens to music. Stirnimann argued that this movement is caused by sensations of hearing in the foetus, but as Moog (trans. 1976) pointed out, Stirnimann's research does not rule out the possibility that it is the mother alone who hears the music, and that the foetus merely reacts to subtle changes in the mother's physiological state.

Moog's own research is extensive, and is almost ethological in approach. Initial observation of his own and friends' children led him to argue that reactions to specifically musical sounds emerge at about the time of a baby's first smile, and that most music has a soothing, rather than an arousing, effect on young babies. He thought this effect was particularly noticeable if the sounds were in the higher frequency range rather than being low-pitched, and that

it held true for both strange and familiar sounds. To try to provide some empirical support for these ideas, and also simply to collect information about the musical experience of a child in the first few years of life, he embarked on a series of studies in which he observed the reactions of children to a specially designed selection of musical stimuli. The first two types of test material were three familiar children's songs and three passages with no melody. In one of the latter the meaning of the words and the rhythmic pattern were of equal importance, in the second words were subordinate to the rhythmic pattern so were not spoken in the rhythm normally demanded, and in the third nonsense syllables were used to make rhythmic patterns. He also included a series of tests consisting of "pure" rhythms (percussion, clapping, and stamping); some instrumental music which included a single note melody, an excerpt from a Bruckner symphony, and a pop song; and a series of "cacophonies" which were generated by having one or more instruments in a string quartet playing in a different key to the others. The series of tests finished with nonmusical sounds.

During his research, Moog carried out over 8,000 individual tests on nearly 500 children, and also evaluated the observations of about 1,000 parents (presumably the parents of the 500 children). He found it difficult to administer the tests to children who were younger than about 5 months old, but suggested that before this age, and after the first few weeks of life, music has a calming effect: as his preliminary observations had led him to believe. However, the

style in which he reported his work occasionally leads to problems in distinguishing empirical findings from anecdotes. For instance, referring to the calming effect of music he said:

All the lullabies which have been handed down for generations show how widely this reaction of babies has been exploited for centuries. The effect is even stronger if the baby is rocked or cradled at the same time; the rocking movement on its own without the singing has a calming effect, but when the two are added together the effect is doubly great. (Moog, trans. 1976, p.50).

It is difficult to tell from the context whether or not this assertion has been experimentally confirmed.

Nevertheless, as a result of playing his series of tests to children, he observed that at the age of about 6 months, instead of being soothed, children displayed a definite interest and pleasure in music. This was usually passive, but could be distinguished from the previous stage in that it might typically involve the cessation of some activity in order to attend to the source of the sound. He found some differences in responses to the different tests, suggesting that children of this age were able to distinguish between, for example, a nursery song and a series of rhythms. In fact the series of nonmusical noises and the series of rhythms aroused very little reaction, although the latter series were the loudest of all the test items. The "cacophonies" attracted



more attention, but the most attention-provoking were the nursery songs and the instrumental music. Moog suggested that this was because of "...the noticeably sweet sound of the two series." (p.55) and argued that at this early stage a baby attends to neither rhythm nor intensity of sound, nor to any words which might be present, but to the quality of the sound itself. In his own words, "...the child does not respond indiscriminately to any musical sound, but...he responds by selecting the sensuously beautiful sound. Thus even at six months old baby is moved by the beauty of sound." (p.55). There are those who would have qualms as to whether "baby" has a conception of beauty which corresponds to that of Moog, some years his or her senior; but the important point to be made here is that Moog's observations led him to believe that children respond to sound quality, or timbre, before they respond to rhythm or loudness.

After the age of passive interest Moog identified a stage in which children start to respond actively to music, for instance by swaying from side to side or bouncing up and down. This may happen any time between the ages of 5 and 8 months. Moog observed that the rhythm of the movements did not correspond to the rhythm of the music, and, paradoxically, very few rhythmic motor responses were produced by the test which consisted of "pure rhythm" (hand-clapping or percussion alone). Moog argued that the movements were in response to both the sound quality and the rhythm, and that they were "...related not only to the sound at a given moment, but also to the sequence in a short passage of music. So they must be

taken as the beginning of a response to rhythm in music." (p.58).

Although children of less than a year old may respond to music by making sounds of their own, Moog was convinced that at this age children are not able to distinguish differences in pitch or in melodic lines. However, between the ages of 1 and 2, about half the children he tested were able to copy differences in pitch. With the development of linguistic ability, words became successively more important. Moog noticed that 2- to 3-year-olds were better at reproducing the words of a song than they were at copying the rhythm and pitch, and that it was not necessarily the fact that the words had meaning that made them the easiest element to grasp. Words presented rhythmically, and nonsense words spoken in a rhythm, were perceived as music (insofar as they produced rhythmic movement and singing, rather than speech, in the children), whereas "pure rhythms" were ignored by 37% of the children even when their volume was considerably increased. Moog found that only 14% of the children regarded the test series of pure rhythms as music, and argued that this was because of the absence of words. He also remarked upon the difficulty that 3-year-olds had in identifying a song when they heard it as a melody alone, without words.

The age of 3 years or thereabouts emerged as being important to Moog: he wrote,

Until the age of about three the development of a response to music was determined by innate ability and by

the musical stimuli presented to the child more or less haphazardly....Far and away the most important factor in determining the difference in response was the difference in musical or general ability." (p.113).

He continued by suggesting that, at the age of 3 years, the determinants of the direction of musical development change from the genetic to the environmental, and he supported this by citing the advantages that children of 3 years and older who are taught many songs and musical games at home have over those who are not so familiar with music. He also pointed out that, earlier than this age, children who were constantly exposed to music (according to their parents' reports) showed little difference in response from children who heard only "...carefully weighed amounts of carefully chosen musical stimuli." (p.113) and that this was quite the opposite effect to the one he had expected.

Moog's studies led him to believe that, after the age of 3 or 4 years, musical development "...goes on peacefully and calmly..." (p.123). He emphasised the importance of training, but said unequivocally, "The child is deaf to harmony at least up till the end of his sixth year, and probably for a long time after that." (p.136). His rationale for this statement was that no children showed any sign of displeasure at the "cacophonies" in the test series, of which he said, "It is hard to imagine a more unpleasant collection of dissonances....Anyone with any capacity for hearing harmony would show some sign of displeasure." (p.26). However, he

reported earlier that children under a year old made fewer responses to the "cacophonies" than they did to otherwise comparable instrumental music, which suggests that, even if they showed no displeasure, they could at least perceive a difference between the former and a stimulus which Moog might describe as "harmonious".

Although Moog's work might be criticised for its rather anecdotal quality, and for his tendency to attribute to very young children sensations which may not be appropriate, it nevertheless makes an important contribution to research in the development of music perception. In its observational, almost ethological approach it provides an example of a solution to one of the main problems associated with the study of the development of musical understanding, appreciation, and ability: that of the measurement of any response a child may make. There is a double difficulty: (a) music is a nonverbal medium, and responses to it may be emotional or physiological or otherwise unobvious; (b) even if older subjects can be persuaded to verbalise their responses, this is impossible with infants, and can be a strategy which leads to confusion (of both the experimenter and the child!) in the case of younger children. As McMahon (1982) pointed out:

Our understanding of children's perceptions of sounds and the development of concepts is restricted by the difficulty of differentiating between the existence of a concept as such and the "possession of a vocabulary" with which to express that concept (Zimmerman, 1975). The

dimensions of this problem are compounded by the frequent mismatch between the words used by adults and the meaning attributed to them by children.... (p.82).

Difficulty in applying adjectives (such as "high" and "low") to music in an unambiguous way is not restricted to young children: to take one example, it is the author's experience that many adults find the distinction between "high" and "low" strings on a guitar confusing, because they are not sure whether the spatial and musical meanings of the words correspond. In this case they do not, because the highest string, musically, is the lowest, spatially. Musicians may fail to clarify this when instructing untrained adults, having assumed that the relevant word will be interpreted according to its musical meaning. However, as Crowther and Durkin (1982) pointed out, adults normally find little problem in understanding the musical meaning of terms that are also spatial, such as "up/down", "high/low", "rising/falling", and "long/short": in fact they regard the musical use of such terms as part of the "core meaning" of the term, rather than as a metaphorical application of an exclusively spatial term. This may not necessarily be the case where children are concerned: in an experiment by Hargreaves, Castell and Crowther (unpublished) it was observed during data collection that when 6- to 8-year-olds were asked to say whether two melodies differed when one was played in a higher key than the other, the adjectives they used were very varied, ranging from "high/low" in some 8-year-olds, through "light/dark" to some



unique contrasting pairs such as "happy/gropey" in younger children. Many children were unable to express how the melodies differed beyond saying, for example, "It's a different note".

Despite some conflicting reports in the literature which may be a result of problems arising when children are asked to give verbal responses, McMahon (1982) argued on the basis of evidence presented at the Fourth International Seminar on Research in Music Education (1974) that an infant's first discrimination among sounds is of timbre, then of loudness, and then of pitch. This argument concurs with Moog's findings described previously: he believed that sensitivity to timbre emerged earlier than sensitivity to rhythm, loudness or pitch. However, the situation is not as clear-cut as might appear at first sight because there is evidence that sensitivity to one musical parameter (e.g. timbre) does not necessarily precede sensitivity to another in a simple linear fashion, but that they are interdependent. For instance, Hermanson (1971) found significant differences in the pitch acuity of kindergarten children according to the timbre of the musical stimulus: responses to a woman's voice were significantly better than responses to a piano or oscillator. Sims et al. (1982) found that 5- and 6-year-old children performed better on a pitch-matching task when the stimulus was a female, rather than a male, voice. They suggested that it was the stimulus range, rather than the timbre per se, which accounted for this difference, owing to the difficulty children of this age have in echoing sounds which are heard in a register other than

that in which their own voices lie. This study indicates that even in a situation where a nonverbal measure of pitch sensitivity is used there are factors other than the age of a child which should be taken into account when studying the development of a specific musical ability.

To return to the problems associated with the interpretation of verbal responses, Andrews and Madeira's (1977) research suggests that many young children may fail to complete pitch discrimination tests not as the result of an inability to detect pitch differences, but because of an inability to deal with the relational language used in such tasks. Moffit (1971), and Trehub and Rabinovitch (1972), also provide evidence that children can detect minimal differences in sounds, even though they may not possess the linguistic ability to describe these. McMahon's (1982) findings were that out of twenty, 5-year old subjects, ten failed to correctly identify changes in pitch when loudness was held constant, or changes in loudness when pitch was held constant. However, half the incorrect responses named changes in pitch in terms of loudness, although level of loudness was held constant: therefore the subjects were aware of some change, but could not identify it correctly.

It is difficult to see any clear way around the problem of determining whether a given "inappropriate" response to a musical task on the part of a child is due to (a) the child's inability to express his/her thoughts in an unambiguous way, although s/he knows the correct response; or (b) the fact that the child actually has not understood the task and arrived at

an answer, regardless of whether or not s/he could express it if s/he had. Although it may not be possible to solve this problem, it is certainly possible to bear it in mind at all stages of research, from designing an experiment to interpreting its results; and it is also possible to look for alternative ways of asking any given question so that conclusions as to children's musical abilities can be validated by several independent approaches. Results of studies such as Andrews and Madeira's (1977) and McMahon's (1982) suggest at least that it may not be meaningless to consider separately the verbal expression of a response and the perceptual processes that gave rise to it.

Many researchers have attempted to outline a series of stages of musical development which are passed through in a more or less invariant order as a child gets older. Some research focuses on the attainment of specific abilities such as pitch acuity, sensitivity to rhythm, sensitivity to timbre, and loudness discrimination (Petzold, 1966, 1969; McMahon, 1982; Moog, 1976). In such studies, the emphasis is on the observable manifestations of a child's level of development, with little reference to the psychological component of these changes in ability. For instance, Moog (1976) gave a detailed account of what a child can and cannot do, musically, between the ages of 6 months and 6 years, but he spent little time discussing what it is in the cognitive make-up of a 5-year-old which enables him or her to "keep time" to music in a way which cannot be achieved by a 2-year-old; or why only 40% of 4-year-olds are able to recognise a familiar song when its



tune is played without the words, whereas 75% of 5-year-olds have no difficulty with this task (Moog, 1976, p.135). Sims et al. (1982) pointed out that Petzold (1966, 1969) suggested that distinct singing ability levels exist as a function of age and that, "He indicates significant improvement in children's accuracy and control of melodic singing between the first and second grades." (Sims et al, 1982, p.104); but they did not discuss the nature of any psychological changes which might accompany or give rise to the improvement.

A slightly different emphasis appears in the work of Bamberger (1975), Machotka (1966), Pflederer (1964, 1967), and Gardner (1973). Although all were interested in the development of musical ability, their interest centred on the cognitive or other processes via which a child reaches a response, rather than on the description of responses per se. All four authors referred to the work of Piaget, and it is in this respect that they have something in common: all of them discuss the growth of musical intelligence in terms of the existence of a series of qualitatively distinct stages of comprehension, and they assess the extent to which these stages correspond to the pre-operational, concrete operational, and formal operational stages described by Piaget. For instance, Pflederer, (1964, 1967); and Pflederer Zimmerman and Sechrest (1970) examined the development of musical concepts in children from a Piagetian point of view by introducing the idea of musical conservation, which they define as "...the ability of an individual to retain the idea of sameness with regard to a complex musical stimulus despite

deformations of temporarily irrelevant aspects of that stimulus." (Pfleiderer Zimmerman and Sechrest, 1970, p.25). Examples of the types of deformations that were used by these investigators are a change of rhythm, a change of instrument, a change of harmony, a change of contour, a change of interval, a change of mode, and a change of tempo. A typical conservation task would involve the presentation of a pair of stimuli, consisting of a standard and a deformation of that standard: in any given deformation, only one of the parameters mentioned above would be altered, and all other variables would remain unchanged. The questions put to the child, on hearing the stimuli, were designed to match Piaget's method of questioning as closely as possible, and were as follows:

You will hear this short tune played several times. (Play the stimulus.) Each time that it is played you will hear it followed by a second tune. I would like for you to answer these questions about the two tunes: Is the second tune the same or is it different from the first tune? (Then, according to the subject's answer): What was different about it? Are they the same in any way? How do you know? Tell me more about what you heard. (Pfleiderer Zimmerman and Sechrest, 1970, p.27).

The child's response, in order to be counted as "correct", had to be that the two tunes were both the same and different: a child who answered simply "the same" or "different" was demonstrating non-conservation. It was found that age was a

significant factor in success on the task; also that an improvement in conservation of tonal patterns preceded conservation of rhythm; and that changes of mode, contour, and rhythm interfered with conservation more than changes in instrument, tempo, or the addition of harmony.

Thus the results of Pflederer Zimmerman and Sechrest's (1970) research do add support to the notion that the concept of conservation may be a valid explanatory construct in the development of musical intelligence. However, as Serafine (1980) pointed out, it may not be necessary to invoke such a construct:

...while the discovery of the relationship between task performance and age is a crucial one, it is not sufficient evidence for either the validity of the tasks or the hypothesizing of stages. Indeed, in almost any cognitive or perceptual task, older children can be expected to do better than younger ones. (Serafine, 1980, p.9).

So although it has been found that performance on "Piagetian" music tasks improves with age (Pflederer Zimmerman and Sechrest, 1970; Jones, 1976; King, 1972), it is perhaps premature to conclude that music conservation tasks are exactly equivalent to Piaget's conservation tasks, or even that an explanation of musical development in terms of conservation is necessarily the most economical or elegant available.

Most of the "Piagetian" music research has been based upon conservation experiments (Serafine, 1980), but there are instances where studies have investigated other aspects of Piaget's theory: for example, Larsen (1973) developed a four-part melodic variation task, parts of which were assumed to require the use of formal operational (as opposed to concrete operational) structures for their successful completion. Bamberger (1975) investigated the way in which children who had improvised a simple rhythmic figure represented this when asked to, "Draw a picture of your claps so you can remember them next week or so someone else could play them". (p.1) She used these representations as a basis for further work in which children of varying ages were asked to make drawings of rhythms, and both children and adults were asked to perform (by clapping) from the drawings already made. She distinguished two contrasting representational strategies, and the first of these she described as the figural strategy. This is closely related to body movement and may be thought of as collecting rhythmic events into chunks each of which reflects a given movement, and which she called the individual's "felt path" through a series of actions (introduction, p.ii). The second strategy she described as the metric strategy, because its focus is the measurement of durations relative to a fixed reference point. She said,

While metric strategy...provides a single schema for classifying events, it is not responsive to context; events which are the same in duration remain the same in

spite of contextual function and regardless of the position of the event in a particular figure.

(introduction, p.ii).

She pointed out that the figural and metric strategies may be associated with pre-operational, in contrast to more operational, behaviour respectively; but also that although the figural strategy does characterise the behaviour of children under 7 years old, it is not necessarily limited to this age group and may extend even into adult thinking. However, her results showed that any adults who did use the figural strategy were quite readily able to comprehend and apply the metric strategy once it had been explained to them, and this was not true of the pre-operational children in her sample. Only older children (10 to 12 years) and adults spontaneously made use of a fully realised metric strategy. In addition to age, musical background appeared to be important in determining what type of strategy was adopted. She found the following:

...metric strategy is characteristic of those who play a musical instrument and who regularly read music from a score. In contrast, performers who "play by ear" spontaneously use a figural strategy as do individuals who do not play an instrument but said they "knew how to read music". Finally, it seems that the two strategies most often function separately even among those who have access to both. (Bamberger, 1975, introduction, p.iii).



Thus Bamberger was able to identify qualitatively different stages in development, insofar as pre-operational children were not able to adopt a metric strategy even after it was explained to them, whereas adults and older children were able to do this. To this extent her findings may be subsumed under a Piagetian framework, but she herself did not lay claims to any theoretical breakthroughs: her primary concern was that the existence of both representational strategies should be recognised in education so that they can both be used, and so that the possibility is acknowledged of developing "...ways of teaching which will encourage children to move freely back and forth between these ways of knowing, each enriching the other." (Bamberger, 1975, p.33).

Machotka (1966) examined changes between the ages of 6 and 12 years in the criteria which boys used to evaluate works of art (in this case, paintings rather than music), in order to assess the extent to which developmental changes mirror the progression of intellectual functioning described by Piaget. He presented reproductions of paintings in groups of three to each child individually, and for each triad the child was asked which painting he liked most and which he liked least, and to give reasons for each of these choices. He argued that justifications referring to subject matter and colour would precede those referring to realism, contrast and harmony of colours, and clarity of presentation, because the former would require no more than pre-operational thought; whereas in order to judge a painting as realistic a child would have to compare the picture in front of him with what he imagined the

appearance of the real object to be. This latter process, Machotka suggested, would require mental processes characteristic of the operational stage, and would therefore not appear until after the age of about 7 years.

These predictions were confirmed by his results. He was also interested in the development of an awareness of style in paintings, and reported that style and composition as criteria of judgement first appeared close to the beginning of adolescence. He proposed that this phenomenon may be a reflection of the changes in intellectual processes which take place with formal operational thought, and that the emergence of formal operational thought may also explain the decrease he observed in the importance of realism at the age of 11 years:

Formal thought means the capacity to reason about verbal propositions; the propositions themselves need (sic) not be true but may remain as mere hypotheses. The criteria of style and of composition appear to imply the hypothetical existence of several manners of representation, one of which (the style or composition that the observer is commenting on at the moment) seems the most satisfactory. The observer cannot judge style or composition if he knows only one....The child who has not reached the hypothetico-deductive level of thought should have difficulty in using these two criteria. When he has reached that level he may begin to lose his interest in realism...his ability to conceive of several styles at once may weaken the attractiveness of the one that he was

capable of distinguishing up until then. (Machotka, 1966, pp. 883-884).

It could be argued that Machotka's findings, particularly as regards style sensitivity, might also apply in the field of musical development. Gardner (1973) reported an experiment in which he devised a way of studying children's sensitivity to musical styles. Previous research of his own (Gardner, 1970 and 1971) had led him to believe that children at or before the concrete operational level tend to focus on the subject matter or dominant figure in a work of art, but that 7-year-old children could be trained to sort paintings according to style, even if competing figural cues were present. In order to test the extent to which similar tendencies might exist with respect to music appreciation, he operationalised style sensitivity in music as the ability to judge whether or not two fragments of music came from the same composition. Musical figure was defined as a solo voice against an instrumental background. The music used was drawn from the repertoire of Western classical music written between 1680 and 1960, and this was presented to each subject in pairs of excerpts, each pair consisting of two halves which were 15 seconds long. Each subject heard 16 pairs of excerpts altogether, of which 8 pairs consisted of two excerpts from the same composition (so 8 compositions were represented). The other 8 pairs consisted of two excerpts from different compositions by different composers (thus a further 16 compositions were represented). Within each group of 8 pairs,

4 pairs were exclusively instrumental and the remaining 4 were mixed, featuring a soloist singing with the orchestra in only the first or second excerpt of the pair. Gardner hypothesised that younger subjects might assume that members of a mixed pair could not have come from the same composition. Four eras of composition were represented, and it was hypothesised that subjects "...would be more likely to recognise pieces composed in widely separated eras as from different composers than pieces composed in the same era." (pp. 69-70). Subjects were 10 males and 10 females at each of five age levels (modal ages 6, 8, 11, 14 and 18-19 respectively), and Gardner pointed out that they were "...overwhelmingly middle class and of high intelligence." (p. 68). On each presentation of a pair of excerpts, a subject was asked whether or not he thought the two members of the pair had come from the same piece of music. A prize was offered for achieving a certain score.

Results indicated that style sensitivity improved with age from 6 through to 11 years, with the 11-, 14-, and 18- to 19-year-olds performing at a very similar level. However, Gardner expressed surprise at the overall excellence of subjects at the task: all age groups performed rather better than he had anticipated. He tentatively attributed this to the gifted population from which they were drawn. One of the most interesting aspects of his findings is that the criteria children used to judge stylistic similarity changed qualitatively rather than quantitatively with age. Thus 6-year-olds could give very little justification for their responses, 8-year-olds tended to use metaphors from outside



music, while 11-year-olds judged from an objective viewpoint. This latter age-group explicitly acknowledged that two discontinuous excerpts could come from the same composition, and this contrasted sharply with the behaviour of the 6-year-olds, for whom excerpts had to sound nearly identical or directly continuous before they were judged as being from the same composition. The oldest groups of subjects (14 and 18 to 19 years) were very aware of the concept of musical style and spoke in terms of musical categories rather than specific features of each stimulus. This could on occasion lead to errors of judgement: Gardner cited responses to two excerpts from the same piece in which subjects thought the excerpts might be "Bach and Handel" or "Mozart and Haydn": thus the older subjects' very sophistication could be a source of error. As Gardner wrote, "This excessive introspectiveness may be one reason why, using an alternative approach, less sophisticated eleven-year-olds did not perform significantly worse than undergraduates." (p.74). Gardner also pointed out that figural cues in music may not be as clear-cut in their effect on perception of style as they are when subjects are asked to judge paintings. In the latter case figural cues may compete with stylistic cues, particularly where children have not yet reached the concrete operational level. Gardner found that the tendency to focus attention on the figure was less pronounced in the 1973 experiment than in studies of sensitivity to painting styles, and suggested that a vocal line is perhaps not the only aspect of music which can be perceived as "figure".



Gardner's research constitutes a unique and important approach to the study of the development of music perception. There has been very little work on any aspect of style sensitivity, particularly from a developmental point of view, one of the exceptions being Hargreaves (1982) who found that children's tendency to classify music in terms of a stylistic label increases with age. This finding is compatible with Gardner's results, although the context of the experimental task was different. (In Hargreaves' study, children were asked simply to say how two pieces sounded the same or different). Greer, Dorow and Randall (1974) and LeBlanc (1979, 1981) looked at children's preferences for different musical styles, but without examining the mechanisms by which children distinguished among styles. Gardner pointed out that although "...high intelligence and operational thinking undoubtedly contribute to style sensitivity, the particular subject matter being investigated should not be overlooked." (p.76). This issue is discussed further in Part 2.2, and provides some of the rationale for the choice of stimulus material used in the first two experiments in this thesis.

To conclude, it may be said that much of the research into the development of aesthetic reactions to music has been heavily influenced by Piaget's ideas. Apart from the examples given here, Parsons (1976) gives an account of this area of development which is based on Piaget's and Kohlberg's approaches. Gardner (1979), however, has reservations as to the extent to which a Piagetian model of development may be appropriate so far as the arts are concerned. He suggests that

an approach is called for which "...builds upon certain of Piaget's assumptions and methods but which takes into account the specific characteristics of diverse symbol systems and media ...." (p. 73). He contends that Piaget's view of the end-point of cognitive development is a rather narrow one: "In Piaget's view, mature cognition is no less, and no more, than the domain of logical-rational thought: accordingly, his end state is the competent scientist." (p.76). In addition, he thinks that Piaget's disregard of the particular materials or symbol system used to present a problem to a child and to obtain a response constitutes the omission of a potentially fruitful area of investigation. He says, "Where Piaget has concentrated exclusively on the logical-rational thought of the scientist, future investigators are likely to probe the skills needed by radically different kinds of thinkers." (p.77). It remains to be seen whether future investigators will actually take up this approach.

## 2.2: The Present Research

Experiments 1 and 2 reported in this thesis follow on from Gardner's (1973) research which was described in Part 2.1. The rationale and procedure are basically similar to those of Gardner, but some important refinements have been introduced, some of which were suggested either directly or indirectly by Gardner himself in his report of his work.

As Gardner pointed out, the subject matter being investigated may be influential in determining the results of tests of musical (or other) style sensitivity, and with this in mind the stimulus material selected for the present research differed in two respects from that of Gardner. In view of the ambiguity surrounding the notion of "figure" and "ground" as it applies to music, none of the compositions used had a vocal line, and all used the same number of instruments. This was done because it was felt that children do not necessarily perceive a vocal line as a figure: for instance, in an opera performed in a foreign language, where the style of singing may be unfamiliar to a child and the words incomprehensible, it is quite conceivable that a child is no more likely to attend primarily to the vocal line than s/he is to attend to any other aspect of the music. In addition, it could be argued that even if such a vocal line might constitute "figure" to an orchestral "ground", this relationship is not the same as that between, for instance, a

folk song "figure" (where all the words may be easily heard) and a guitar accompaniment "ground". In summary, it was felt that this issue was so complex that it was best omitted from the present studies, the existence of an extra, unknown variable being more likely to confound than to help interpretation of the results.

The second respect in which the stimulus material differed from that of Gardner was with respect to the styles of music represented. It is noticeable, when reading the literature on music psychology, that a very large proportion of the research using naturalistic as opposed to experimental music deals with responses to Western, classical music only: although there are some recent exceptions to this generalisation (e.g. Geringer 1980; Schuckert and McDonald, 1968; Hargreaves and Colman, 1981). However, overall, it seems that styles akin to jazz, rock and pop have been dramatically under-represented. This is perhaps particularly relevant where developmental research is concerned because it is to younger age groups that these styles of music often have special significance (Johnstone and Katz, 1957; Inglefield, 1972; Murdock and Phelps, 1972). This omission is found not only in psychological research but also in education, becoming especially apparent in secondary schools as Burnett (1977) pointed out. Moog (1968, translated 1976) also made a stand against the exclusion of popular music from serious consideration by educators and researchers: he said, "Music educators seem more united in their stand against pop music than on any other matter, but the fact that it is not taught

in school is no reason for not including it in this series of tests." (p.23). This contrasts sharply with the attitude of the author of another musical ability test writing in the same year:

Jazz music was not included, as this would be unlikely to yield examples of really good harmony, would be likely to prejudice the authorities against the tests, and would waste the children's time if they were listening to poor music. (Wing, 1968, p.37)

In the light of the discussion above and other indications of the importance of "popular" music to a majority of the population (not least the fact that it is named "popular" music; and also that so far as radio, television, newspapers, and other media are concerned, considerably more coverage appears to be given to "popular" than to "serious" music) the stimulus material for the present research was selected so that both "popular" and "classical" styles were represented equally. It should be pointed out that both these stylistic categories are used here in a very broad sense, "popular" music including jazz, rock, and many other varieties of "non-serious" music; "classical" referring to all Western art music written between about 1500 and the present day. Throughout the remainder of this thesis it can be assumed that these are the senses in which the two categories are to be understood unless otherwise indicated; and in future they will be written without quotation marks.



Gardner (1973) suggested that the surprisingly high scores achieved by his subjects may have been due to the population from which the subjects were drawn, because the subjects involved were all of high intelligence and came from white, middle-class backgrounds. These features, like the inclusion of classical music only, may limit the generality of his findings; and for this reason the subjects in the present studies came from a wide range of socioeconomic and ethnic backgrounds. Gardner thought that a further contributing factor in the achievement of a large proportion of correct responses might have been the existence of a reward: in the present studies, no reward was offered as it was felt that this would give a closer approximation to a "real-life" situation.

The remaining modifications of Gardner's procedure are concerned with his operationalisation of style sensitivity. Two problems emerged here. He found that his youngest subjects (4- to 6-year-olds) required excerpts to sound almost identical before they would say that they were from the same piece. Although this could be a reflection of a rather inflexible approach to the concept of style, it could also be due to a misinterpretation of the use of the words "same" and "different" in this context: in other words, the youngest subjects' low scores could be a result of linguistic difficulties rather than (or as well as) a lower level of musical ability than the older subjects. In order to minimise any possible linguistic confusion in the present experiment, the task was presented in a story-like context, and a

secondary operationalisation of style sensitivity was used: i.e. in addition to being asked whether or not they thought two excerpts of music might have come from the same piece, subjects were asked whether or not the excerpts might have been written by the same composer. It was anticipated that subjects might adopt rather less strict similarity criteria for the "same composer" judgements than for the "same composition" judgements, because the "composer" question was the less specific of the two.

The second problem with Gardner's operationalisation, as it stands, was that his oldest subjects (14- to 18-year-olds) tended to possess an implicit awareness of the concept of style and therefore realised that two excerpts which were in fact from the same piece could have had different sources, because they knew it was possible for two different compositions to be of the same style. This type of mistake arose specifically because they were aware of style, and although their answers reflected a high level of awareness of the concept they were construed as errors. To provide a check on the extent to which this happened in the present studies, a record was kept of whether each error made was of Type II (i.e. saying excerpts were from the same source when they were actually from different sources) or of Type I (i.e. saying that excerpts were from different sources when they were actually from the same source), so that the age distribution of the types of error could be examined.

As in Gardner's experiment, subjects were asked to provide justifications of their initial responses as to

whether excerpts came from the same or different compositions. The criteria upon which they based their judgements were then examined, and in the present case this was accomplished by applying to these criteria a system of content analysis derived by Hargreaves and Colman (1981) from several musical and aesthetic theories. Hargreaves and Colman found that responses to music could usually be subsumed under one of five categories, which may briefly be described as follows: (a) Categorical responses, which classify music in terms of a stylistic label such as "jazz", "blues", or "Baroque"; (b) Objective-Analytic responses, which refer to intrinsic qualities of the music such as instrumentation or tempo; (c) Objective-Global responses, which although they also refer to intrinsic qualities of the music, describe the character of the music as a whole rather than specific aspects of it; (d) Affective responses, which involve an emotional and subjective evaluation of music; and (e) Associative responses, which refer to extramusical associations evoked by the sound of the music. An attempt was made in the present experiments to assimilate all subjects' justifications under these five categories: the extent to which this proved possible will be discussed later.

The aims of Experiments 1 and 2 may be summarised as follows:

1. The sensitivity to musical style of subjects aged between 7 and 18 years was explored, using two operationalisations of style sensitivity, with particular reference to the questions of whether sensitivity varied with

age, and/or with the nature of the stimulus material (in this case responses to classical music were compared with those to popular music).

2. The nature of errors made by subjects in judging whether two excerpts of music had the same source or different sources was examined. The investigation centred upon the age distribution of Type I and Type II errors, and also the distribution across types of stimulus material (classical, or popular, pairs of excerpts) of these errors.

3. The age distribution of different criteria used by subjects to justify their responses was examined. On the basis of Gardner's (1973) findings, it was felt that younger subjects might be more likely to give Affective or Associative responses than Analytic ones, using Hargreaves and Colman's (1981) system of content analysis, whereas the reverse might be true for older subjects.

## 2.3: Experiment 1

### Sensitivity to Stylistic Differences in Classical and Popular Music: 8- to 11-Year-Olds

#### Method

##### Subjects.

The subjects were 120 children from two junior schools in Leicester. There were two groups aged 8 to 9 years and 10 to 11 years respectively, with 60 children in each group. Within each group there were equal proportions of children from each of the two schools, and roughly equal numbers of males and females. The schools had very different catchment areas, one serving part of the town in which many "disadvantaged" families had been housed, the other having a mainly upper middle class intake.

##### Musical material.

This was chosen on the basis of two pilot studies so that all compositions were, as far as possible, equally familiar to the subjects and were matched with respect to volume and the number of instruments involved. None contained a vocal line.

Excerpts were taken from six compositions, three of those



compositions being popular and three classical. (Popular and classical are used here to designate very broad stylistic areas, as was pointed out in Part 2.2.) The three classical compositions were (A) (Anonymous): Four Chansons, "The Art of the Recorder", David Munrow; (B) Mozart: String Quartet K458, first movement; and (C) Ravel: String Quartet in F major, first movement. The three popular compositions were (a) Modern Jazz Quartet: Festival Sketch, Stockholm Concert; (b) Focus: Carnival Fugue; and (c) Pink Floyd: Meddle, Echoes.

Four, 15-second excerpts were taken from each composition. The stimuli to be presented consisted of pairs of these excerpts, members of a pair being separated by a silence lasting 5 seconds. Six pairs of excerpts were exclusively classical, and six were exclusively popular: that is, in no case were both styles of music represented within one pair. In each set of six pairs, three had members both of which were from the same composition, and three had members each of which was from a different composition. Four excerpts were taken from each composition in order to avoid using any excerpt more than once, and in the case of the pairs where excerpts came from the same composition, the second member of a pair was never directly continuous with the first.

To summarise the position: if the three classical compositions are represented by uppercase letters A, B and C, and the three popular compositions by lowercase letters a, b and c; and if the suffixes 1, 2, 3, and 4 are used to indicate different excerpts from the same composition, then the excerpts taken were as follows: A1,A2,A3,A4; B1,B2,B3,B4;

C1,C2, C3,C4; a1,a2,a3,a4; b1,b2,b3,b4; and c1,c2,c3,c4. These were combined in the following manner to give 12 pairs, of four types: (1) classical, same: A1-A2, B1-B2, C1-C2; (2) classical, different: A3-B3, A4-C3, B4-C4; (3) popular, same: a1-a2, b1-b2, c1-c2; (3) popular, different: a3-b3, a4-c3, b4-c4.

Excerpts were recorded from records onto a master tape, and then onto two cassette tapes so that the 12 pairs appeared in a different random order on each tape (but maintaining the same order within pairs). On each of the final cassette tapes the same "practice" pair (taken from compositions other than those used as experimental material) preceded the 12 pairs of excerpts. Details of the practice pair and of the random orders of presentation are shown in Appendix 2.3.1.

### Questionnaires.

Data were collected by asking each subject to give written responses to the following questions after s/he had heard each pair of excerpts:

- (a) "Do you think that they are from the same piece?"
- (b) "If not, could the same person have composed them?"
- (c) "Have you heard either before?"
- (d) "What made you decide whether or not they were from the same piece?"

These questions were presented in a booklet consisting of three pages, each page containing four sets of the four

questions. A separate slip of paper presenting the same four questions was used for the "practice" pair. A sample page from a booklet appears in Appendix 2.3.2.

#### Procedure.

Subjects were tested in groups of 10 to 28, according to their availability, with half the groups hearing one order of presentation of the excerpts and the other half hearing the second order. Testing sessions were arranged so as to ensure that each order of presentation was eventually heard by equal numbers of subjects. At the beginning of each testing session subjects were told:

I'd like you to imagine you're in a large house with lots of rooms. You come to a door and open it, and find a room full of all sorts of musicians and musical instruments. You listen to what they're playing for a moment, like this, (first excerpt of "practice" pair is played on the cassette recorder) and then you go away for a while: it might be for a few seconds or a few hours. When you come back they are still playing, like this, (second excerpt of "practice" pair is played). What I'd like you to tell me is whether or not the musicians are still playing part of the same piece. I shall give you a booklet to write your answers in. You might decide that even if the excerpts of music do not sound as if they come from the same piece, they could still have been written by the same person. If so, you can tell me that by your answer

to the second question of each set in your booklet. I would also like you to say whether you've heard the pieces before, and if you can, explain what made you decide whether or not they were from the same piece of music.

At this point each subject was given a slip of paper on which were printed the questions (a) to (d) quoted previously. After ensuring that each subject thoroughly understood the instructions by inspecting their responses and, if necessary, repeating the "practice" pair, response booklets were distributed and the 12 pairs of excerpts were played. The tape was stopped after each pair and enough time was allowed between pairs to let the slowest subject write his or her answers. Each testing session took about 25 minutes. Before returning their booklets, subjects were required to write down their age and sex on the front cover. Each booklet was marked on reception according to which of the two random orders had been presented.

#### Analysis of Responses and Results

Responses to question (a) were given a score of 1 if the subject correctly identified the excerpts as from the same or from different compositions. No score was given for an incorrect response, but a record was kept of whether the subject had made the mistake of saying that two excerpts came from the same composition when in fact they had different

sources (Type II error), or the opposite mistake or saying that two excerpts came from different compositions when they actually had the same source (Type I error).

A Pearson's product-moment correlation coefficient was computed between the total accuracy scores for each stimulus pair for each of the two tapes, over all subjects, to find out whether order of presentation might have affected subjects' accuracy. A high positive correlation was found ( $r = 0.87$ ,  $p < 0.001$ ) so it was assumed that order of presentation was not a confounding variable and results from the two tapes were pooled.

Three separate analyses were carried out, the first two concerning style sensitivity (incorporating responses to questions [a] and [b]), and the third to ascertain the degree to which subjects were familiar with the excerpts (question [c]). The criteria upon which subjects had based their judgements were examined, but not subjected to any statistical test (question [d]).

#### First and second analyses: style sensitivity.

Each subject was given four scores determined by the number of correct responses given to question (a) for each of the four types of pair (i.e. classical, same; classical, different; popular, same; popular, different). Because there were three of each type of pair the range of possible scores was from 0 to 3 for each type.

Before analysing these data, a similar procedure was carried out with respect to question (b). It was found that a



number of subjects had misunderstood the task, because they had answered "Yes" to question (a) and "No" to question (b), implying that they thought that excerpts were from the same composition but had been written by different composers! Data from these subjects were discarded. After balancing numbers in the two age groups by making random withdrawals, the final number of subjects whose data was included in the first, third and fourth analyses was 106, there being 53 in each age group. Some subjects consistently omitted question (b), so the second analysis was based on data from 86 subjects.

Two, 3-way analyses of variance were performed on responses to questions (a) and (b) respectively (age x style of music x type of pair). "Type of pair" was a repeated measures factor, because all subjects heard both "same" and "different" pairs; and "style of music" was also a repeated measures factor: all subjects heard both classical and popular pairs.

Two significant main effects emerged from the first analysis (based on responses to question [a]), one for age ( $F [1,104] = 6.75, p < 0.05$ ), and one for type of music ( $F [1,104] = 106.26, p < 0.001$ ). A summary table for the analysis appears in Table 2.3.1. Examination of the means revealed that the younger age group performed with greater accuracy than the older group (means were 2.42 and 2.28 respectively), and that subjects gave more correct responses for popular than for classical excerpts (means were 2.62 and 2.08 respectively).

There were also three significant 2-way interactions. The

first was between age and style of music ( $F [1,104] = 6.41$ ,  $p < 0.05$ ), the second was between age and type of pair ( $F [1,104] = 4.01$ ,  $p < 0.05$ ), and the third was between type of pair and style of music ( $F [1,104] = 6.69$ ,  $p < 0.05$ ). The 3-way interaction (age x style of music x type of pair) was also significant ( $F [1,104] = 4.28$ ,  $p < 0.05$ ). The significant interaction effects are illustrated graphically in Figures 2.3.1 to 2.3.4.

The second analysis (responses to question [b]) gave rise to three significant main effects, for age ( $F [1,84] = 12.79$ ,  $p < 0.001$ ), for type of music ( $F [1,84] = 25.21$ ,  $p < 0.001$ ) and for type of pair ( $F [1,84] = 141.23$ ,  $p < 0.001$ ). A summary table for this analysis appears in Table 2.3.2. Examination of the means showed that the younger subjects made fewer errors than the older group (overall mean accuracy scores were 2.20 and 1.94 respectively), that more errors were made for classical than for popular pairs (overall mean accuracy scores were 1.92 and 2.22 respectively), and that more errors were made for "different" than for "same" pairs (overall mean accuracy scores were 1.44 and 2.70 respectively). There was in addition one significant 2-way interaction, between age and style of music ( $F [1,84] = 6.78$ ,  $p < 0.025$ ). This is illustrated in Figure 2.3.5. The mean numbers of correct responses given in each condition under each operationalisation of style sensitivity are shown in Figure 2.3.6.

### Third analysis: familiarity.

Subjects were asked in question (c) to indicate whether or not they had heard each excerpt before. Each excerpt was given a score of 1 (heard before) or 0 (not heard before) accordingly. Because each composition was represented by four separate excerpts, familiar compositions should have scored 4 and unfamiliar compositions 0, if subjects were responding accurately and consistently. In practice what happened was that subjects tended to say they had heard some excerpts of a composition before but not others from the same composition, so that each composition emerged with a familiarity rating of between 0 and 4 for each subject. A repeated measures  $t$ -test was performed between familiarity ratings for classical compositions and familiarity ratings for popular compositions, over all subjects, and it was found that  $t = 4.46$ , with 105 degrees of freedom,  $p < 0.001$ . Examination of the means suggested that the three popular compositions were rated as more familiar, overall, than the classical compositions (mean familiarity ratings were 1.26 and 0.91 respectively. Theoretical maximum rating = 4).

### Examination of justifications for responses.

In question (d), subjects were asked to give justifications for their responses to questions (a) and (b). Using some of the criteria outlined by Hargreaves and Colman (1981) it was possible to place subjects' justifications in seven categories, and these are outlined below.

(i) Tempo/Rhythm/Speed. Included in this category were all

responses where a subject had commented on some aspect of the rhythm or tempo of excerpts in a pair as being the basis upon which s/he had made a decision about whether or not the two members of a pair were from the same composition or by the same composer. Examples of responses which would be included in this category are, "The beat was different" and "One was slow one wasn't".

(ii) Instrumentation. Included in this category were all justifications in which a subject had commented upon the instrumentation of excerpts: for example, "The cymbals told me they were different" and "One had a flute in it".

(iii) Timbre/Pitch/Melody. This category included all responses where a subject had commented upon the timbre, pitch or melody of an excerpt. The three qualities were included in one category because it was often not apparent to which one a subject was referring. Examples of responses falling into this category are, "One was high, one low", and "It had a different note".

(iv) Associative/Affective. This category included two slightly different types of response which had in common the feature of not referring to explicit intrinsic characteristics of the music. Associative responses were those which linked the sound of the music to some nonmusical event (e.g. "It reminds me of a moon"), and Affective responses were those which described the emotional or subjective reaction of the listener as being the basis upon which s/he made a response: for example, "One scary, one happy and scary".

(v) Categorical. Included in this category were all



responses which referred specifically to the style of the music, for instance, "It sounded like dance music from the time of Henry VIIIth" and "It was jazz".

(vi) Similarity. This category was typified by the response "It just sounds the same": it included all responses indicating that subjects were aware of similarities or differences between excerpts, and felt that they had more to say than that they did not know why they had made a particular response, but were unable to express what it was that gave rise to this awareness. A second example is "One was different, the other wasn't".

(vii) Other. In this category were included all responses which did not fall into any of the previous six, but were too different to form a further homogeneous category. Several examples of this category are given: "They both rymes" (sic), "They a bit of a different groove" (sic), "One was normal the other wasn't", "The joints were funny, one was noisy and the other quiet", "They are not right", "I herd both befor so I now" (sic), "Because it look very like the first one" (sic), "They had different nose is" (sic), "Becomes I said No" (sic), "Didn't care", "It was rubbish", "Because my brain toed me" (sic).

In theory each subject should have made 12 or more responses (i.e. at least one response to each of 12 pairs of excerpts). In practice, there were many occasions on which subjects answered "I don't know" in response to question (d). These latter responses were not included in the data which is discussed here, but each of the other responses was placed in



one of the seven categories, giving each subject seven scores. These scores had a potential range from 0 to at least 12, although in practice no subject used any one category of response more than six times. Histograms were drawn to show the total number of responses which fell into each category for each style of music, for each age group separately. These histograms are shown in Figure 2.3.7.

### Discussion

In this experiment style sensitivity has been operationalised as the ability to tell (a) whether or not two excerpts of music come from the same composition, and (b) whether or not two excerpts of music are by the same composer. It was anticipated that these abilities might improve with age, particularly as Machotka (1966) has, with reference to the visual arts, argued that the concept of style might emerge only with Piaget's formal operations stage at around the age of 11, because before this age children would be incapable of imagining that it was possible for there to be several different ways (or styles) in which a particular object could be represented. A young child could not judge style, therefore, because one can only do this by comparing the style in question with "...others which, at the time of judgement, are imagined or hypothetical." (Machotka, 1966, p.884). So far as the visual arts are concerned it is at least physically possible to have two paintings of different styles present at the same time. However, it could be argued that Machotka's

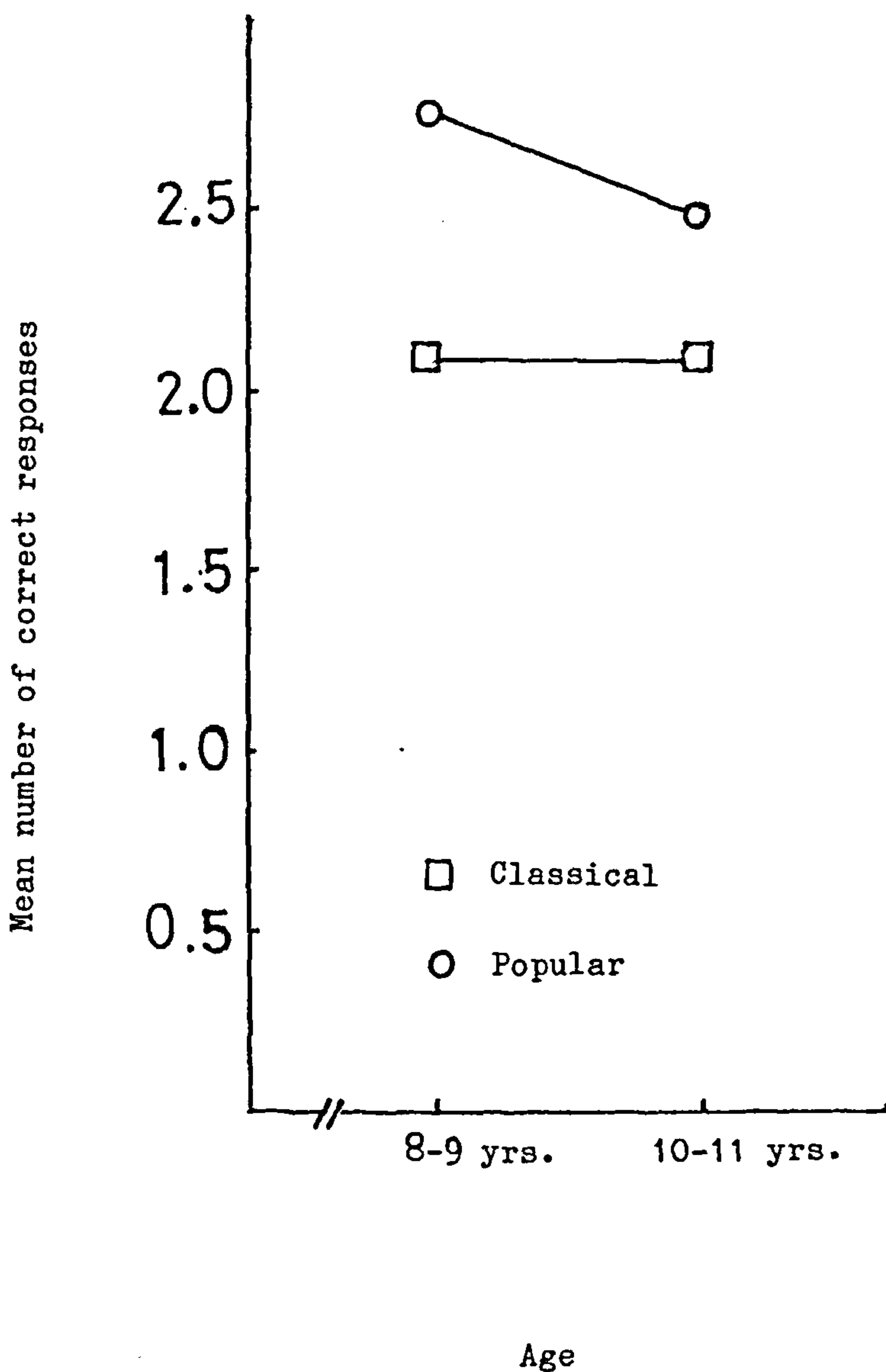


Figure 2.3.1: Significant two-way interaction.  
The effects on style sensitivity (first operationalis-  
ation) of age of subjects, and whether stimulus pairs  
were popular or classical.

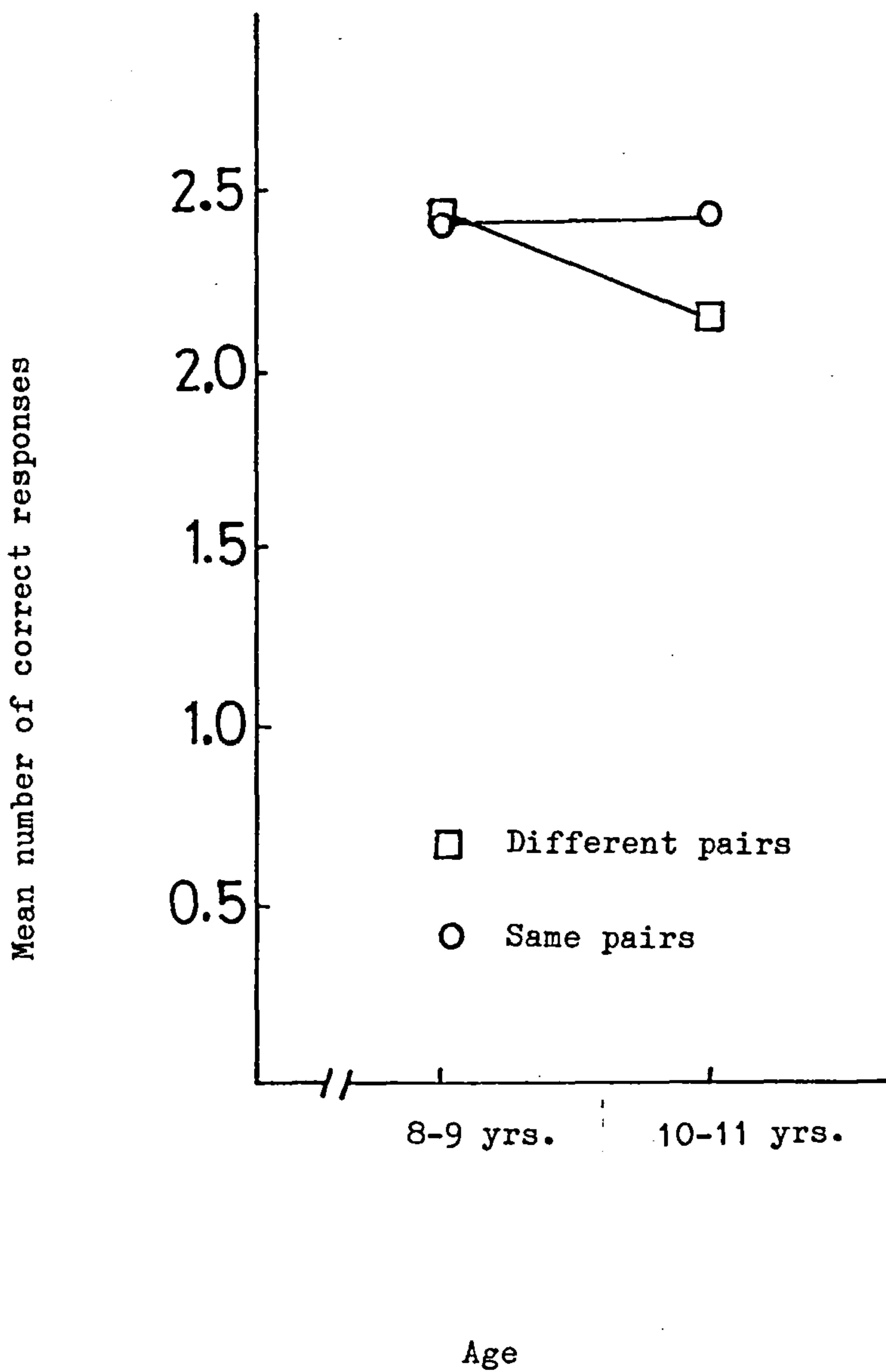


Figure 2.3.2: Significant two-way interaction. The effects on style sensitivity (first operationalisation) of age of subjects, and whether excerpts in a stimulus pair were from the same or different sources.

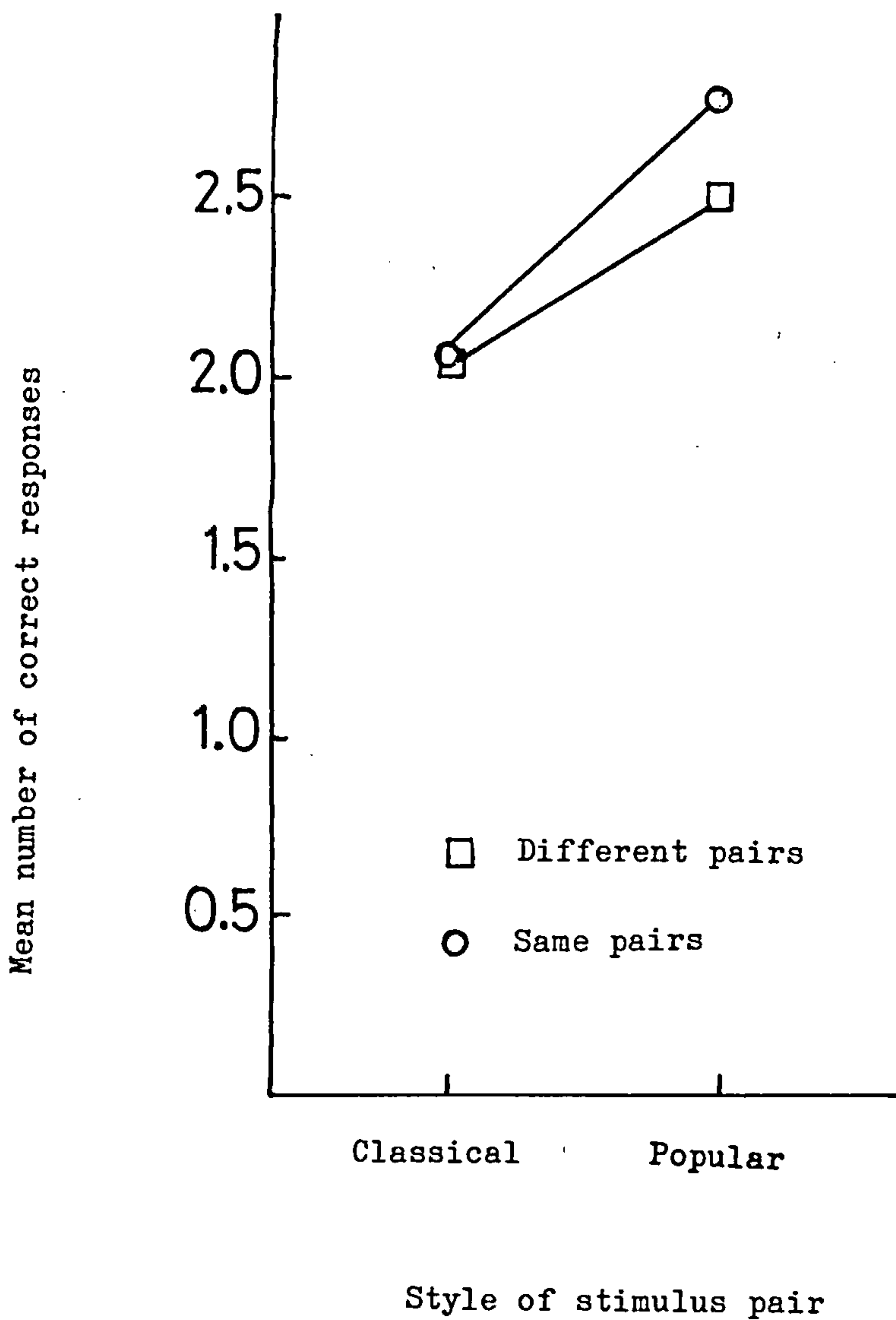


Figure 2.3.3: Significant two-way interaction.  
The effects on style sensitivity (first operationalisation) of style of stimulus pair, and whether excerpts in a pair were from the same or different sources.

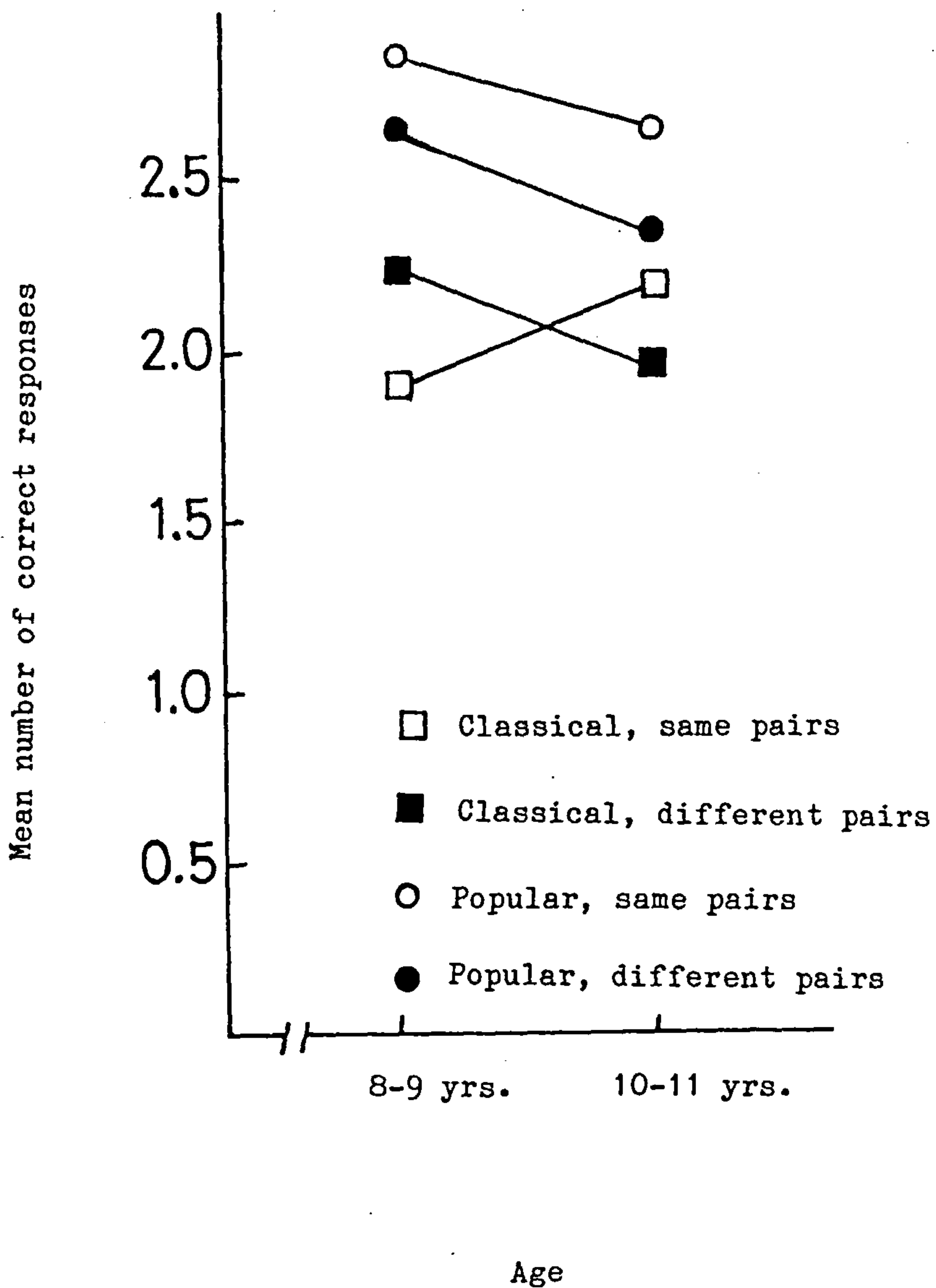


Figure 2.3.4: Significant three-way interaction. The effects on style sensitivity (first operationalisation) of age of subjects, style of stimulus pair, and whether excerpts in a pair were from the same or different sources.



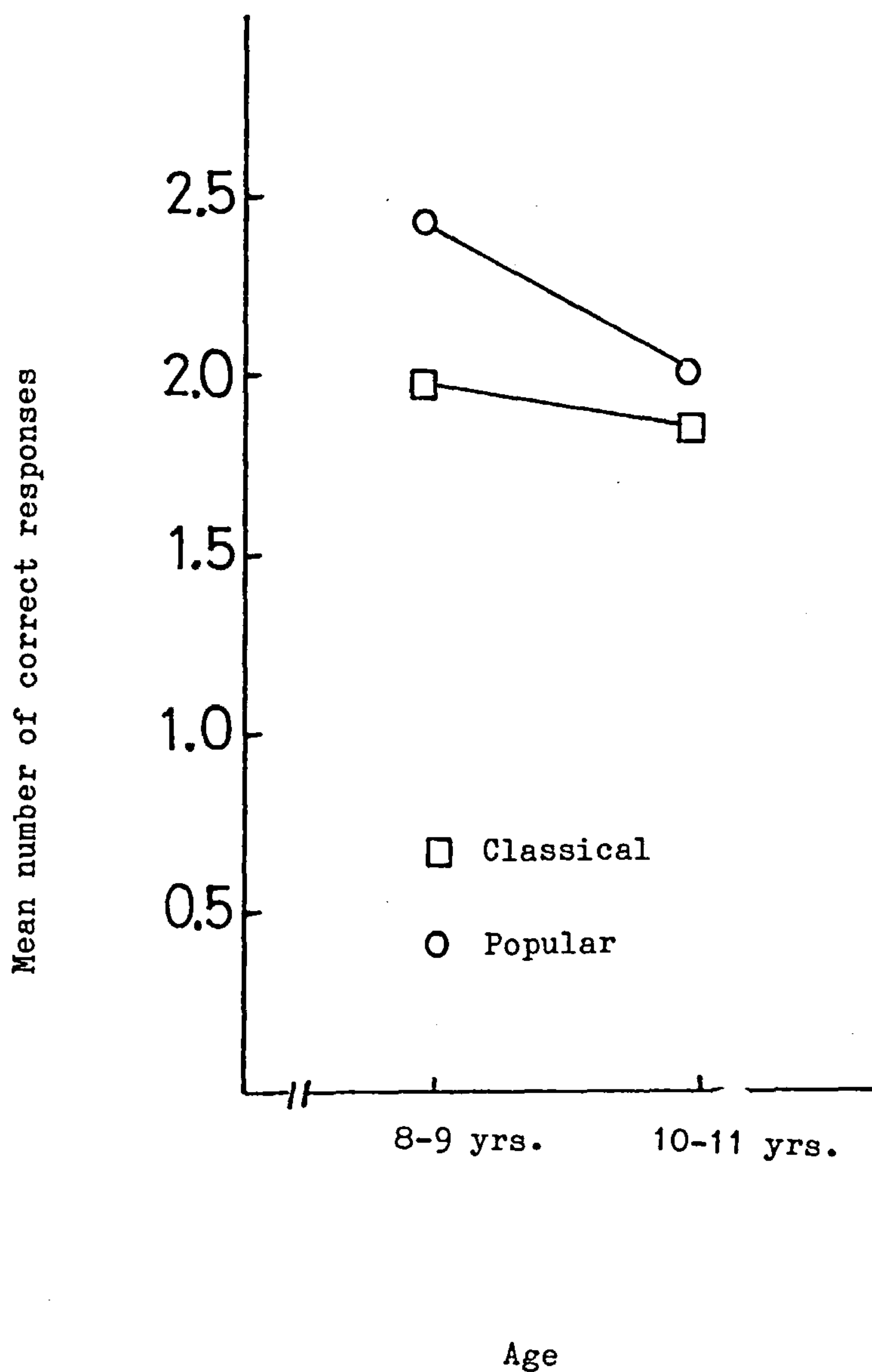


Figure 2.3.5: Significant two-way interaction.  
The effects on style sensitivity (second operational-  
isation) of age of subjects, and style of stimulus  
pair.

Mean number of correct responses

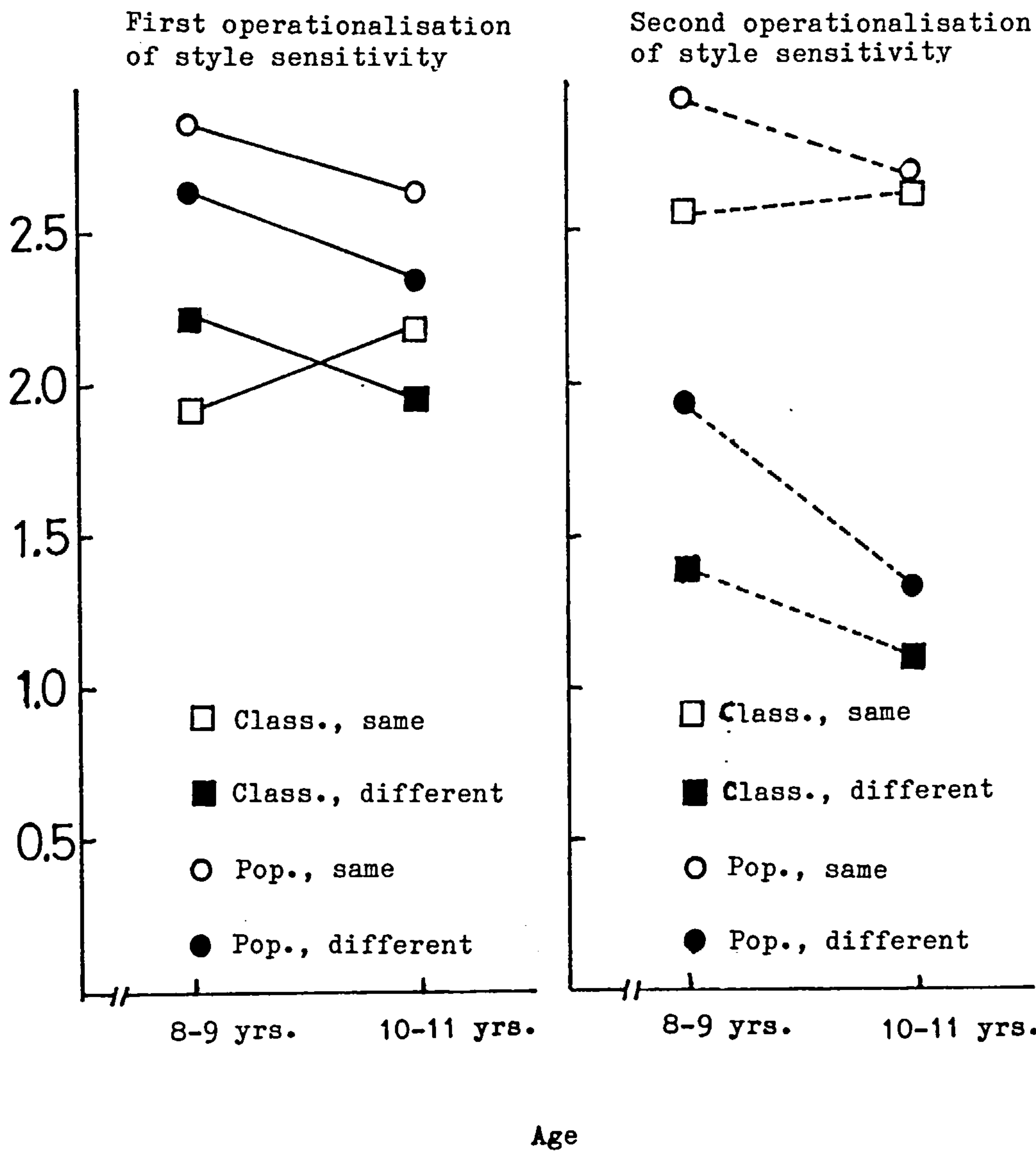


Figure 2.3.6: Amount of style sensitivity displayed, using two operationalisations of the same, in each condition.

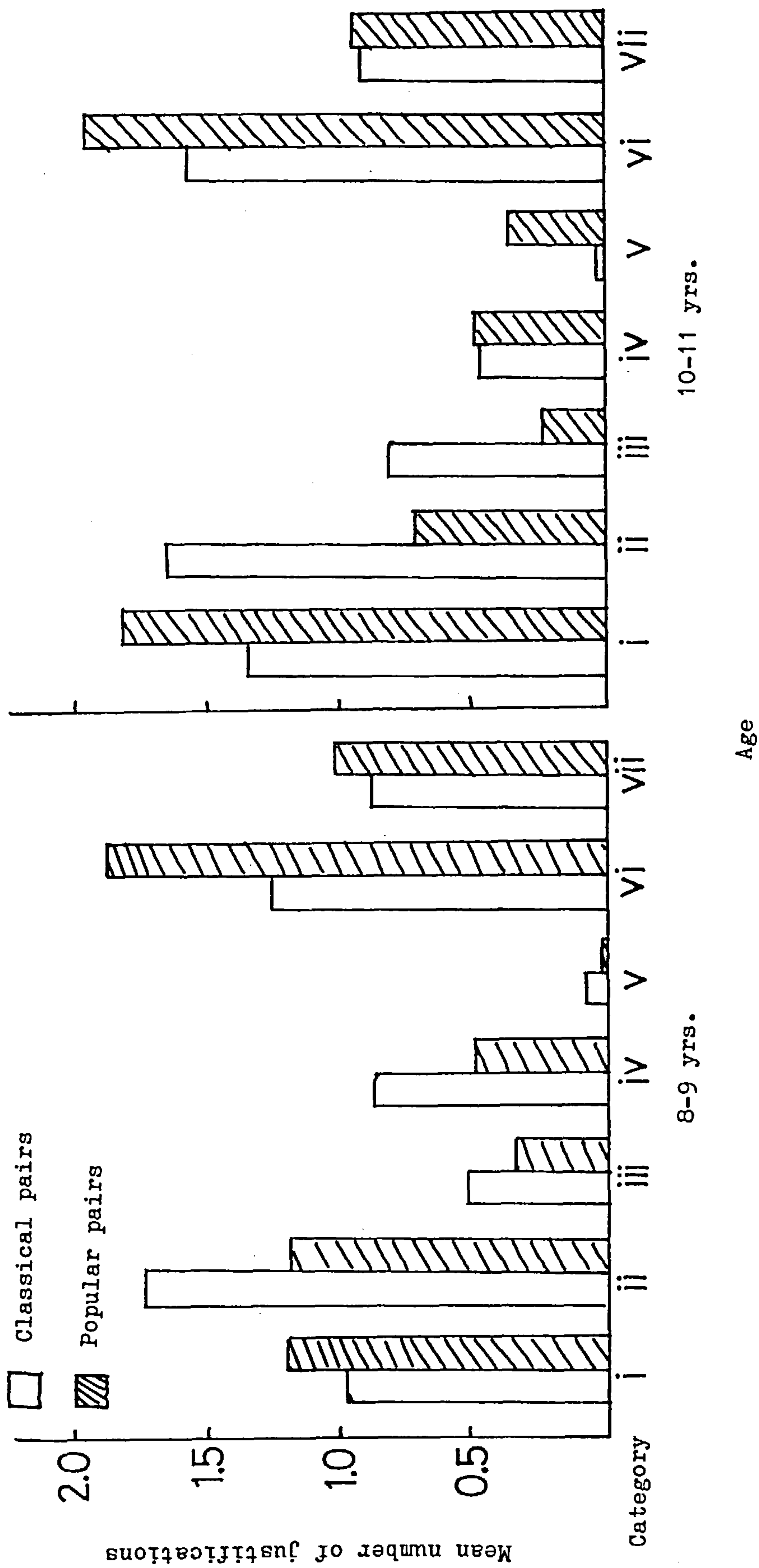


Figure 2.3.7: Numbers of justifications for responses given by each age group falling into each category, for classical and popular stimulus pairs.

Table 2.3.1

Analysis of variance summary table,  $N=106$   
The effects of age, style of music, and type of stimulus pair on style sensitivity (first operationalisation).

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	2.123	6.75	<0.05
ERROR	104	0.315		
B	1	30.651	106.26	<0.001
AB	1	1.849	6.41	<0.05
ERROR	104	0.288		
C	1	1.142	2.57	n.s.
AC	1	2.123	4.01	<0.05
ERROR	104	0.529		
BC	1	2.123	6.69	<0.05
ABC	1	1.594	4.28	<0.05
ERROR	104	0.317		

A=age, 8-9 years and 10-11 years (independent factor).  
B=style of music, classical and popular (repeated measures factor).  
C=type of stimulus pair, "same" and "different" (repeated measures factor).

Note. Style sensitivity was operationalised as the accuracy with which subjects identified excerpts as being from the same or from different compositions.

Table 2.3.2

Analysis of variance summary table, n=86

The effects of age, style of music, and type of stimulus pair on style sensitivity (alternative operationalisation).

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	5.628	12.79	<0.001
ERROR	84	0.44		
B	1	8.477	25.21	<0.001
AB	1	2.279	6.78	<0.025
ERROR	84	0.336		
C	1	140.698	141.23	<0.001
AC	1	2.616	2.63	n.s.
ERROR	84	0.996		
BC	1	0.744	2.48	n.s.
ABC	1	0.012	0.04	n.s.
ERROR	84	0.301		

A=age, 8-9 years and 10-11 years (independent factor).

B=style of music, classical and popular (repeated measures factor).

C=type of stimulus pair, "same" and "different" (repeated measures factor).

Note. Style sensitivity was operationalised as the accuracy with which subjects identified excerpts as being written by the same or by different composers.



suggestions might be particularly relevant to musical stimuli because these are temporal in nature, rendering it impossible to make a comparison in which both stimuli are actually present and distinguishable.

It was therefore surprising to find that the 10- and 11-year-olds were not as accurate as the 8- and 9-year-olds on both measures of style sensitivity. However, it must be borne in mind that although 8- and 9-year-olds made fewer errors than 10- and 11-year-olds in the popular pairs of excerpts, there was no such age difference for the classical pairs: as can be seen from Figures 2.3.1 and 2.3.5. So far as classical music was concerned, Gardner (1973) found that his older subjects tended to make mistakes because they were over-sensitive to stylistic features: they tended to say that two excerpts came from different sources when in fact they did not (this constitutes a Type I error). It was therefore possible that the lower accuracy of the 10- and 11-year-olds was due to their making more Type I errors than the 8- and 9-year-olds, but examination of Figure 2.3.2 shows that this was not the case. If they had made more Type I errors than the 8- and 9-year-olds, they would have had lower scores than the 8- and 9-year-olds for the "same" pairs, but in fact mean scores for the two age groups on this type of pair were nearly identical. It was when scores for "different" pairs were examined that the age difference became most apparent: 10- and 11-year-olds made more Type II errors than the younger group (i.e. they were more likely than the younger group to say that two excerpts came from the same composition when in fact they

did not).

One confounding variable which might have contributed to the poorer performance of the 10- and 11-year-olds was that they were tested in larger groups than the younger subjects; this was necessitated by the timetabling and room space in the schools taking part, and meant that the 8- and 9-year-olds may have been tested in a less formal, more relaxed atmosphere. However, one might have expected this to influence scores for both styles of music, whereas the poorer performance of the older group manifested itself only in the case of popular music. This suggests further lines of research, because it is possible that by the age of 11 children have formed fairly definite likes and dislikes in the field of current pop music, and cease to distinguish very much among compositions and styles which fall outside the range of their chosen favourites. It is likely that the musical samples used in this experiment were not among their favourites, because one of the criteria by which the compositions were selected was that they had to be relatively unfamiliar. The 8- and 9-year-olds, on the other hand, were perhaps more "open-eared", because there might be less social pressure on them to like certain types of music and dismiss others. (This hypothesis is anecdotally supported by the appearance among the older groups' responses of comments such as "too slow" or "it was rubbish", both of which imply a negative attitude towards the excerpts, and their absence among the responses of the 8- and 9-year-olds.) The notion that peer group norms may be an important factor in determining reactions to music among adolescents is supported

by experiments such as those by Inglefield (1972) and Chapman and Williams (1976). Sociological research (e.g. Murdock and Phelps, 1972) suggests that music, during youth, is a very important group identifier: adolescents in general tend to know a great deal about popular music and artists, and to define themselves in those terms.

Overall, the results of the first analysis may be summarised as follows: (a) 10- and 11-year-olds, if they make errors, are more likely to be "overinclusive" than "underinclusive" for both popular and classical excerpts: in other words they exhibit a tendency to make Type II rather than Type I errors, and this tendency manifests itself regardless of whether the stimuli are popular or classical. (b) 8- and 9-year-olds, if they make errors (which they do less often than the 10- and 11-year-olds, so far as popular, but not classical, pairs are concerned), are likely to be "overinclusive" as regards popular excerpts, but "underinclusive" as regards classical excerpts. In other words, they tend to make Type II errors when responding to popular excerpts, and Type I errors when responding to classical excerpts. (c) Both age groups make fewer errors in response to popular than to classical excerpts.

One interpretation of the age difference (i.e. the finding that 10- and 11-year-olds are less accurate as regards popular music, and perhaps more "overinclusive", than the 8- and 9-year-olds) could be that rather than being worse at perceiving differences between excerpts of music, 10- and 11-year-olds, more than the younger groups, are tolerant of a



wide range of musical possibilities existing within one composition or one composer's style. Further research involving a wider age range would be necessary in order to investigate this argument.

The question as to whether or not the same composer might have written both excerpts in a pair was asked in order to provide a secondary operationalisation of style sensitivity. Responses to both questions gave rise to similar results, suggesting that both were probing a similar concept. It had been anticipated that this would be the case, but also that subjects would adopt a higher similarity criterion for "composition" judgements than for "composer" judgements. This supposition gives rise to two predictions. Firstly, for "same" pairs, the less stringent the similarity criterion adopted, the less chance there would be of making an error. Therefore asking whether each excerpt had been written by the same composer would result in subjects giving more correct responses than would asking whether excerpts came from the same composition. Secondly, for "different" pairs, in which both excerpts came from different compositions by different composers, the less stringent the similarity criterion adopted, the more chance there would be of making an error. This would result in greater accuracy when the question referred to "same composition" than to "same composer" for "different" pairs. Examination of the mean accuracy scores attained in response to each question (Figure 2.3.6) suggests that these predictions were borne out, although no statistical tests were performed to ascertain the significance of these

differences because the clarification of this issue was not a primary aim of the experiment.

However, there was one noteworthy difference between classical and popular music with respect to the effects of the alternative operationalisations, although the statistical significance of this difference remained untested. From Figure 2.3.6 it can be seen that, for classical "same" pairs, the "same composer?" question produced more correct responses than the "same composition?" question; but for popular "same" pairs accuracy was equally high regardless of what the question was, for both age groups. This trend might be explained by suggesting that children may be more generally familiar with popular than with classical music. They may therefore feel less confident of their ability to judge differences and similarities within classical than popular music, and might therefore be less willing to answer in the affirmative to the rather specific "same composition?" question in the former than in the latter case. Their responses to the more general "same composer?" question would not be so much affected because it represents a somewhat safer option than the "same piece?" question. In the case of "same" pairs, this lack of confidence with respect to classical and not popular music would lead to the pattern of results shown in Figure 2.3.6.

It was intended that all the compositions used in the present experiment should be equally familiar or unfamiliar to the subjects, and they were selected on the basis of pilot studies with this in mind. It was decided to check that this end had been achieved by asking subjects whether or not they



had heard each excerpt before, and an interesting problem arose from their responses: if a subject were familiar with a given composition it might be expected that he or she would be equally familiar with all excerpts from that composition. However, it was found that most subjects were rather inconsistent in this respect. Many rated one excerpt from a composition as familiar (i.e. they said they had heard it before) and then rated as unfamiliar a second excerpt that was taken from within 20 seconds of the first. Overall, subjects were more likely to rate popular excerpts than classical ones as being familiar; but the finding that they could not consistently rate compositions suggests that they may have been more familiar with the popular style than the classical style as suggested in the preceding paragraph, and were not necessarily familiar with the specific popular compositions used in the present experiment.

With respect to the justifications given by subjects for their responses to questions (a) and (b), it was found that the older subjects produced more justifications overall than the younger subjects, as can be seen from Figure 2.3.7. (The 10- and 11-year-olds produced 715 justifications, 8- and 9-year-olds produced 662). This finding mirrors that of Hargreaves (1982). As was pointed out in Part 1.2, it is very important to bear in mind that age differences of this kind may be a reflection of greater verbal fluency in the older group as well as of any increase in sensitivity to music. Partly for this reason, and partly because of the nature of the data, no tests were made of the statistical significance

of age differences and of differences between popular and classical music in the number and nature of justifications given. Nevertheless, the findings represented in Figure 2.3.7 are interesting in several respects.

Firstly, it can be seen that subjects in both age groups gave more justifications falling into category (vi) when judging popular than classical pairs of excerpts. Category (vi) included all justifications where subjects had not been able to express why they thought excerpts were from the same or different compositions, they simply said "It sounds the same", for instance. It could be argued that this difference between classical and popular excerpts arose because subjects felt the lack of an adequate vocabulary with which to describe popular music, and that this might be a reflection of the tendency to concentrate on classical music in most schools, so far as music teaching is concerned. This hypothesis is marginally supported by the finding that there were more justifications given for decisions about classical than popular pairs of excerpts, overall (696 and 681 respectively), suggesting that subjects were perhaps more articulate with respect to classical music (although this does not imply greater accuracy with respect to classical than popular music, as the findings from the first analysis of variance indicate).

Secondly, in both age groups, it can be seen from Figure 2.3.7 that justifications involved rhythm or tempo (category [i]) more often for popular pairs of excerpts than for classical pairs, whereas the converse was true of justifications involving instrumentation (category [ii]) or

melody (category [iii]). Popular music is typically seen as having a prominent "beat", so the former finding makes sense intuitively: it implies that subjects were more likely to be consciously aware of rhythmic characteristics of the popular than of the classical excerpts. Similarly, it could be argued that instrumentation and melody are more important in classical than in popular music, and that this was reflected in the latter finding.

In common with Hargreaves' (1982) study, there appeared to be no marked increase with age in the use of any one criterion of judgement relative to the other criteria, with the possible exception of Categorical responses (category [v]) to popular music. Hargreaves suggested that older children may use Categorical responses more often than younger children because their sensitivity to stylistic categories of music is greater. However, the present findings are paradoxical because the instances where more Categorical responses were given (i.e. to popular pairs of excerpts) were exactly the instances where the 10- and 11-year-olds made more errors in style sensitivity than the 8- and 9-year-olds as can be seen in Figure 2.3.1. Figure 2.3.4 shows that these errors were likely to be of Type II (the error of saying that two excerpts had the same source when they did not), so this finding could be explained by suggesting that 10- and 11-year-olds' conception of what constitutes a style of music is unrealistically broad. However, it should be borne in mind that the figures upon which this idea is based are very small and that no test of statistical significance was performed.



All these suggestions are best treated as tentative working hypotheses. Further research is needed, where any conclusions drawn about the relative importance of different physical characteristics of music are not dependent upon written self-report alone, because it was clear in the present experiment that many children found it hard to put into words exactly why they made their judgements. (To such an extent, in the case of one 11-year-old boy, that he resorted to drawing pictures in response to question [d]). The results of this experiment give no reason to believe that style sensitivity and verbal fluency are positively correlated: rather the reverse, because the 8- and 9-year-olds were more accurate, but gave fewer written justifications (Figure 2.3.7) than the 10- and 11-year-olds.

Overall, the ability of primary school children to detect stylistic differences (as operationalised in the present experiment) was surprisingly high. Gardner (1973) also commented on this finding, and he partly attributed his subjects' high sensitivity scores to their unrepresentatively high level of intelligence and the fact that they were all of middle class origin. It is interesting that subjects in the present study, who possessed a wide range of abilities and came from a very wide range of socioeconomic and ethnic backgrounds, also performed very well on the task.

## 2.4: Experiment 2

### Sensitivity to Stylistic Differences

#### in Classical and Popular Music:

#### 7- to 18-Year-Olds

The previous experiment demonstrated that 10- to 11-year-olds appeared to have less sensitivity to style than 8- to 9-year-olds under all conditions investigated, except for that where excerpts in a given pair were classical and from the same composition (Figure 2.3.4). Style sensitivity was operationalised as the ability to tell whether or not two excerpts of music came from the same composition, or were by the same composer, and as already pointed out, Machotka (1966) argued that style sensitivity will only begin to emerge at around the age of 11 years. Bearing this in mind, the findings of Experiment 1 are somewhat counter-intuitive. However, the paradox might be resolved in several ways; firstly it is possible that factors other than age may have been affecting "style sensitivity"; for instance, whether the music is classical or popular would appear to be important in the light of the present findings. Secondly, although the procedure of asking subjects whether or not two excerpts of music might have come from the same composition undoubtedly measures some kind of sensitivity, it may be premature and unnecessary to



call it "style" sensitivity, at least in the same sense as that in which Machotka used the word "style". A third possible explanation of the anomalous findings is that style sensitivity might in fact decrease with age, although intuitively this does seem rather unlikely.

It was pointed out earlier that one interpretation of the age difference could be that 10- to 11-year-olds, rather than being worse than 8- to 9-year-olds at perceiving differences between excerpts, might be more aware than the younger group of the fact that even within one composition music can vary considerably. Subjects in the present experiments are required to make decisions about whether or not excerpts sound sufficiently similar to be perceived as originating from the same composition; and in essence this could be considered to be a decision about the breadth of membership of a category: in this case, the "category" being a given composition of music. The relevance of categorisation in this area will be discussed in detail in Part 4, but the implications of the findings of Experiment 1 could be that 10- to 11-year-olds have a broader conception of what constitutes a given musical category than do 8- to 9-year-olds. In order to investigate this hypothesis further, it was felt that it would be useful to collect data from subjects both older and younger than those already used, to make any developmental trends more apparent. In addition, widening the age range would certainly establish whether or not style sensitivity actually showed a consistent decrease with age. Subjects' justifications for their responses were again examined, this time looking

specifically at the way in which the use of categories (i) and (ii) altered with age and with style of music. On the basis of subjects' responses in Experiment 1 it was felt that category (i) justifications (Tempo/Rhythm/Speed) might occur more often for popular than for classical music, and that category (ii) justifications (Instrumentation) might occur more often for classical than for popular music. The way in which category (vi) (Similarity) was used by different age groups and for different styles of music was also examined, because it seemed likely that the younger subjects might be less articulate than the older ones.

## Method

### Subjects.

There were 80 subjects, 20 in each of four groups whose ages were 7 to 8 years, 8 to 9 years, 10 to 11 years and 18 years or more. The three youngest groups were from two schools in Leicestershire. Data for the 8- to 9- and 10- to 11-year-olds were provided by taking 20 subjects' data at random from each age group used in the previous experiment. The adults were mostly students from Leicester University, with an age range of 18 to 35 years, mean age 21.2 years. There were approximately equal numbers of males and females in each group.

### Musical material and questionnaires.

The musical material used was identical to that described in the previous experiment: see Part 2.3 and Appendix 2.3.1 for full details. The same questionnaire was also used (Appendix 2.3.2) but it was felt that the task might be too demanding for the 7- to 8-year-olds in its complete form, so they were asked questions (a) and (d) only, verbally.

### Procedure.

Because this experiment was an extension of the previous one, the procedure followed for the 7- to 8-year-olds and adults was identical, as far as possible, to that employed in the first study. The younger subjects were tested in groups of 10, and the adults were mainly in smaller groups averaging 4 members. This was unavoidable because the latter group were volunteers and were tested mainly on the University campus when they were available. As in the previous experiment, half the subjects heard one order of presentation of the excerpts and the other half heard the other order.

The youngest group were not presented with the usual booklet of questions because it was felt that they might find the task too difficult; instead each subject was given a blank sheet of paper and questions (a) and (d) only, quoted below, were presented orally by the experimenter each time the tape was stopped between pairs of excerpts. These subjects were given the number of each pair (as determined by its position on the tape) and were asked to write this down before writing their responses to the two questions, in order to avoid any



confusion in scoring the sheets.

In summary, each subject heard 12 pairs of excerpts of music preceded by a "practice" pair. Of these 12 pairs, 6 were classical and 6 were popular. After hearing each pair, all subjects except for those in the youngest group were required to answer four questions about the excerpts in it:

(a) Do you think that they are from the same piece?

(b) If not, could the same person have composed them?

(c) Have you heard either before?

(d) What made you decide whether or not they were from the same piece?

Members of the youngest group were required to answer questions (a) and (d) only.

#### Analysis of Responses and Results

As before, responses from the two experimental tapes were combined for each age group. Responses to question (a) were scored as they were in the first study, so that each subject received four scores ranging from 0 to 3, determined by the number of correct responses given for each of four types of pair (ie classical, same; classical, different; popular, same; and popular, different). A three-way analysis of variance with repeated measures on two factors was performed (age x style of music x type of stimulus pair). Table 2.4.1 contains a summary of the ANOVA. There were two significant main effects, for age ( $F[3,76] = 10.91, p < 0.001$ ) and for style of music ( $F[1,76] = 46.18, p < 0.001$ ). The main effect for age is illustrated in

Figure 2.4.1. So far as the main effect for style of music was concerned, the mean number of correct responses were: for classical, mean = 2.18; for popular, mean = 2.56. There were also two statistically significant two-way interactions, firstly between age and style of music ( $F[3,76] = 6.60$ ,  $p < 0.001$ ) and secondly between age and type of stimulus pair ( $F[3,76] = 3.94$ ,  $p < 0.05$ ). These interactions are illustrated in Figures 2.4.2 and 2.4.3 respectively.

The reasons subjects gave for their responses were also inspected, and each reason was included in one of seven categories as described in the first experiment. However, criteria for inclusion in category (vi) (Similarity) were broadened to encompass responses such as "I don't know" or "not sure" and cases in which subjects had given no reasons at all for their responses. In the present experiment, therefore, category (vi) contained all the "nonspecific" justifications given by subjects. The frequencies with which reasons fell into each category, for classical and popular pairs, are shown in Table 2.4.2. Because all reasons were included in this categorisation and each subject produced about 12 reasons, allocation of reasons to categories was not on a totally independent basis. However, taking any individual category, the number of times it was used by a subject could be considered as a score for that subject, measuring his/her tendency to give that particular type of reason for his/her responses. Using these scores, two analyses of variance were performed on data from selected categories. The first investigated the relationship between age, style of music and



the use of categories (i) and (ii) only: i.e. those mentioning either tempo, rhythm and speed; or instrumentation. This was a three factor analysis of variance with repeated measures on two factors, namely style of music and category. The second analysis was an analysis of variance with repeated measures on one factor (style of music), which took into account reasons falling into category (vi) (nonspecific reasons) only and investigated how the use of this type of reason altered with age and with style of music.

Results of the two analyses are summarised in Tables 2.4.3 and 2.4.4 respectively. The first gave rise to three statistically significant main effects, for age ( $F[3,76] = 19.37$ ,  $p < 0.001$ ), for style of music ( $F[1,76] = 102.54$ ,  $p < 0.001$ ) and for category into which reasons fell ( $F[1,76] = 29.72$ ,  $p < 0.001$ ). So far as the first main effect was concerned, the mean numbers of reasons in categories (i) and (ii) given per subject per style at each age level were, 7- to 8-year-olds: 0.11; 8- to 9-year-olds: 1.21; 10- to 11-year-olds: 1.34; 18 years and over: 2.15. Overall, taking both categories of response together, more reasons were given for decisions about classical than about popular pairs of excerpts (means were 1.21 and 1.20 respectively). More responses fell into category (ii) (Instrumentation) than category (i) (Tempo/Rhythm/Speed): the means were 1.28 and 1.12 respectively. There was in addition one significant two-way interaction (style of music x category,  $F[1,76] = 124.83$ ,  $p < 0.001$ ), and the three-way interaction (age x style of music x category) was also significant: ( $F[3,76] = 10.44$ ,

$p < 0.001$ ). The interactions are illustrated in Figures 2.4.4 and 2.4.5.

The second analysis gave rise to two significant main effects, the first for age, with  $F(3,76) = 42.62$ ,  $p < 0.001$ , the second for style of music:  $F(1,76) = 138.28$ ,  $p < 0.001$ . The interaction between these two factors was not statistically significant. The means for the four age groups, from youngest to oldest, were 5.45, 2.15, 2.10, and 0.53 respectively. Examination of the means revealed that, over all age groups, there were more "nonspecific" reasons given in response to popular than to classical excerpts (means were 2.59 and 2.52 respectively).

## Discussion

It was suggested at the beginning of Part 2.4 that one possible explanation of the findings of the previous experiment was that "style sensitivity", as operationalised here, actually decreases with age. It is obvious from Figure 2.4.1 that this is not the case: the figure shows that adults gave more correct responses than the other three age groups, with the 7- to 8-year-olds making the most errors. Nevertheless the mean score of the sample of 8- to 9-year-olds was still somewhat higher than that of the 10- to 11-year-olds: this is a reflection of the fact that these subjects were a representative sample of the total number tested in the previous experiment.

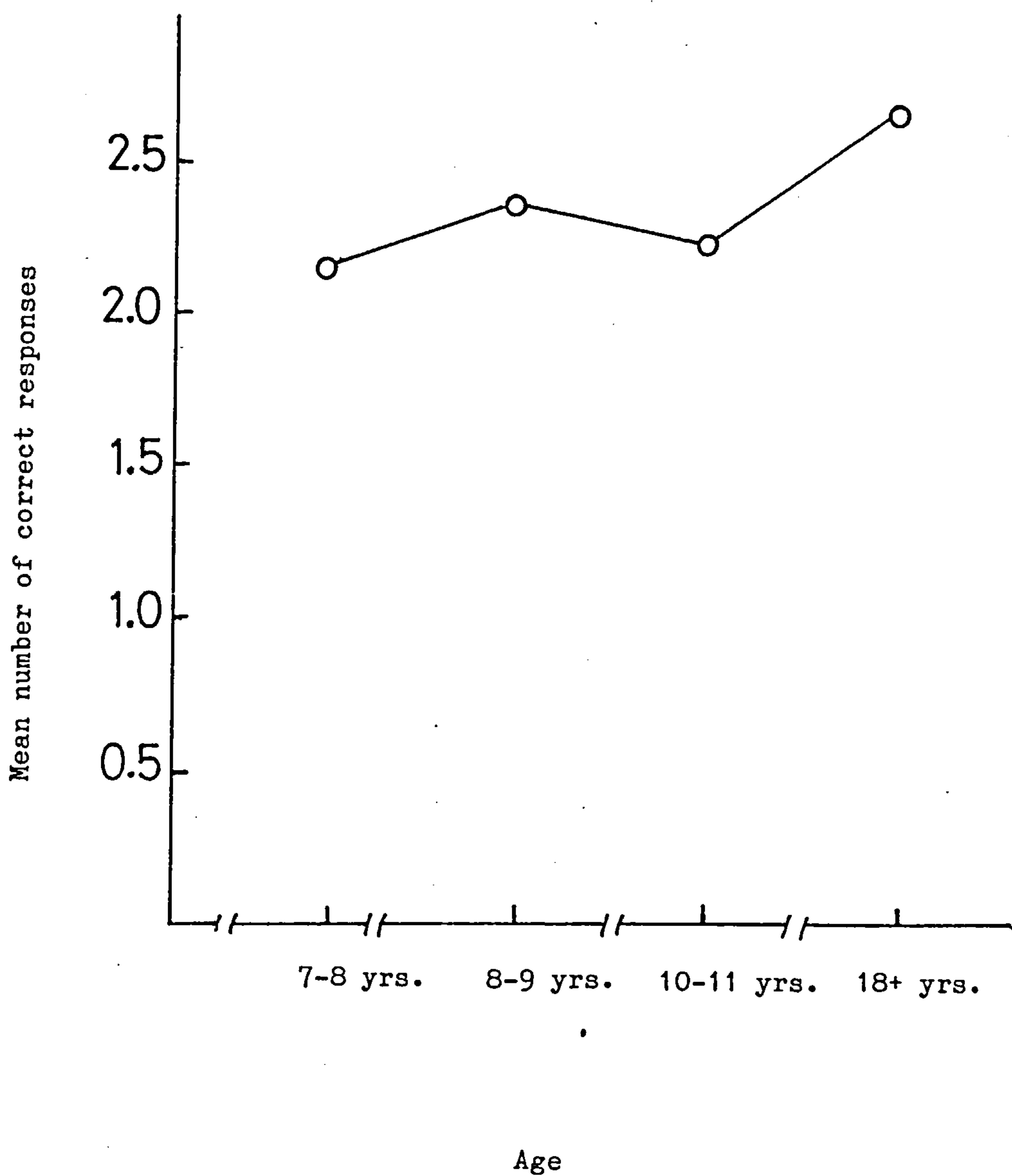


Figure 2.4.1: Significant main effect.  
The effect of age of subjects on style sensitivity  
(first operationalisation).

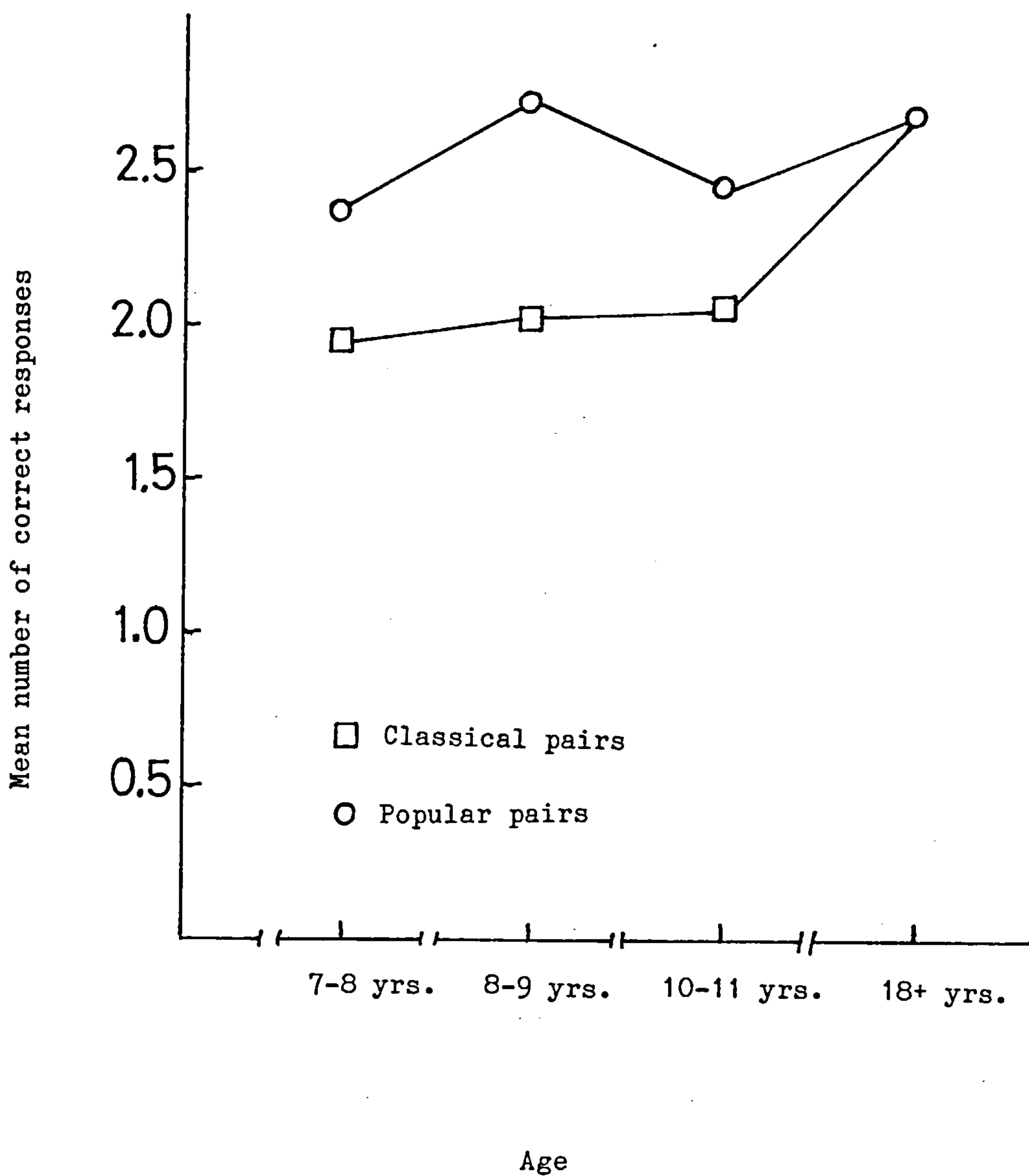


Figure 2.4.2: Significant two-way interaction.  
The effects on style sensitivity (first operationalisation) of age of subjects, and whether stimulus pairs were popular or classical.

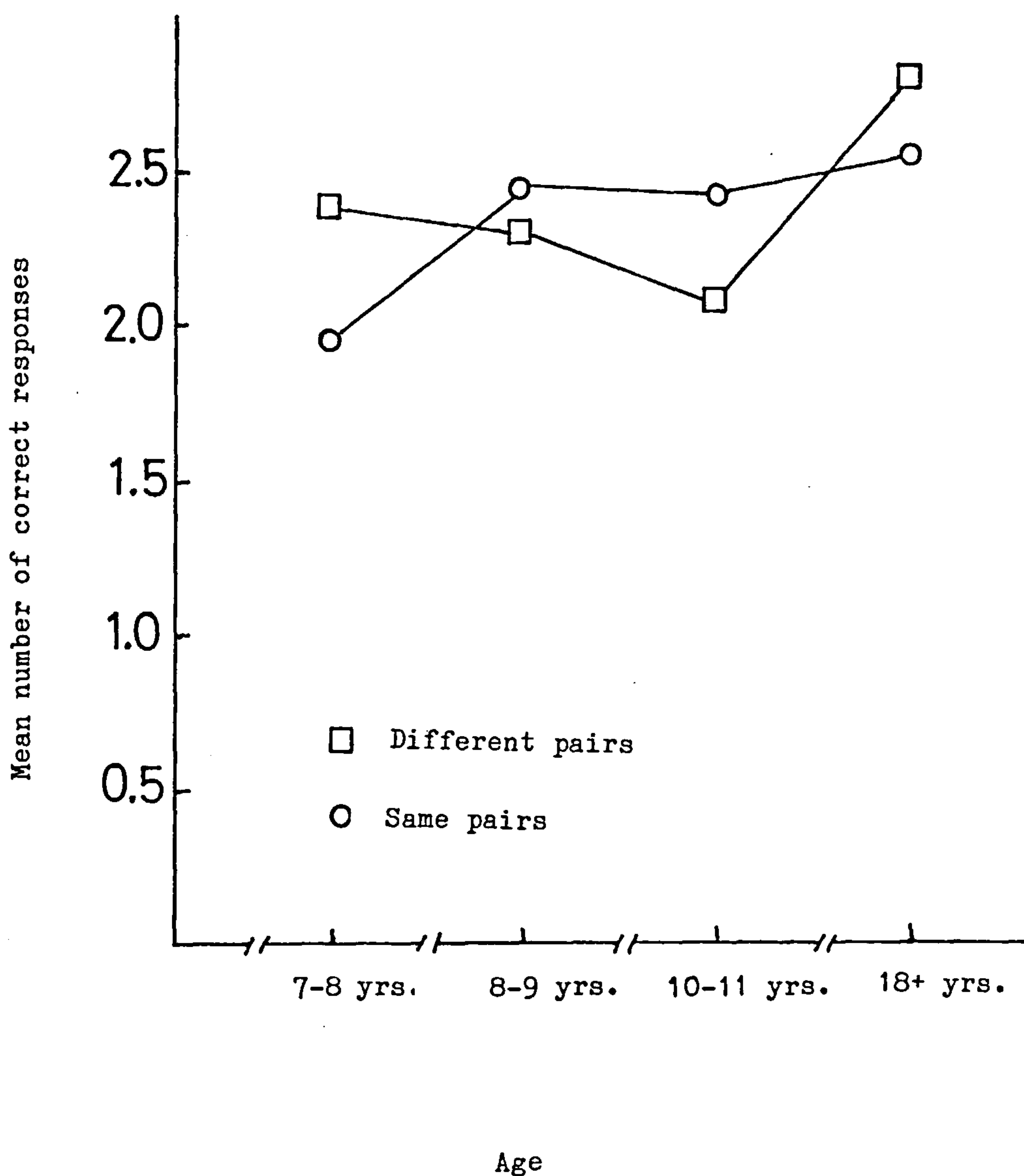


Figure 2.4.3: Significant two-way interaction. The effects on style sensitivity (first operationalisation) of age of subjects, and whether excerpts in a pair were from the same or different sources.



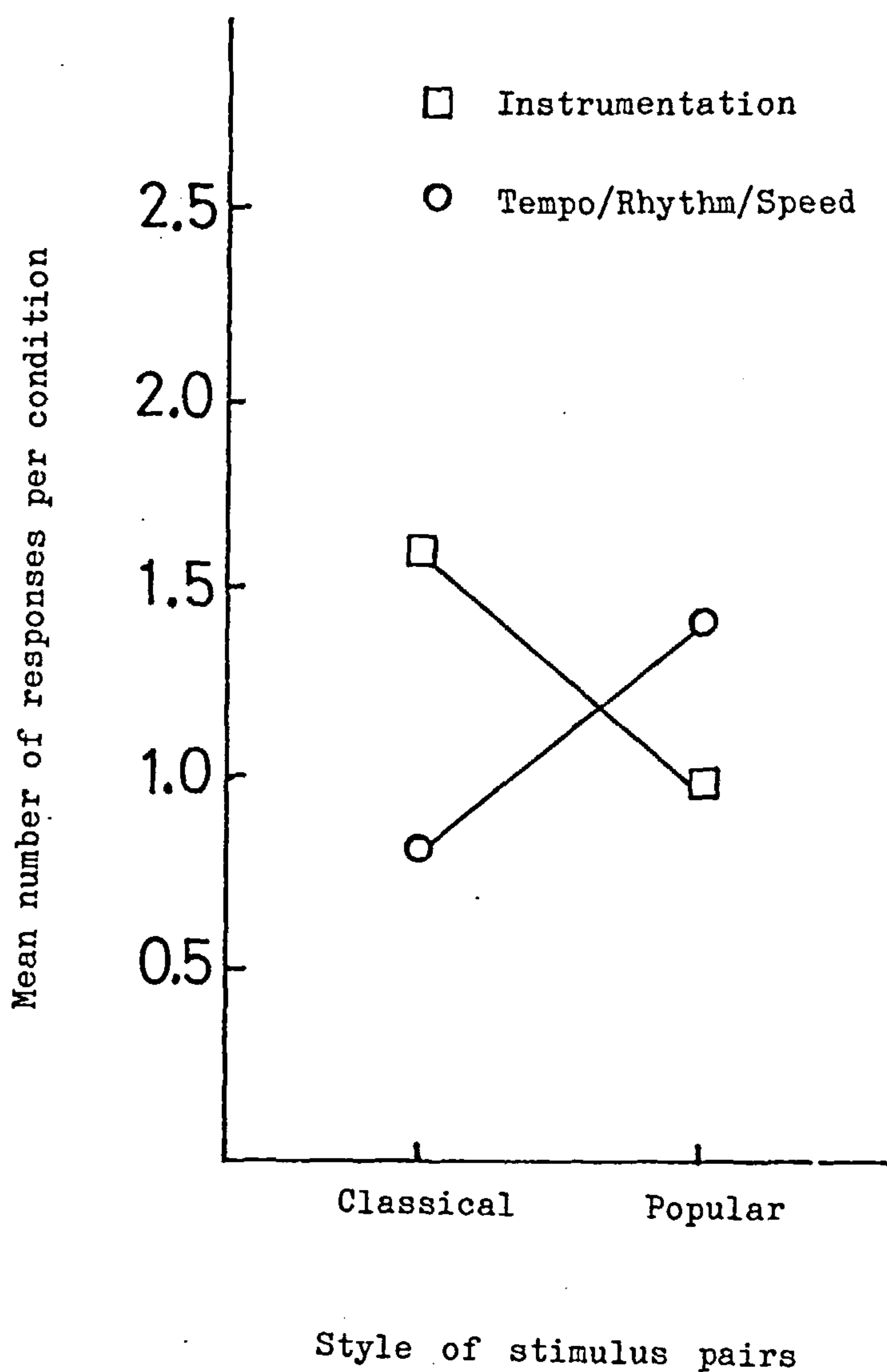


Figure 2.4.4: Significant two-way interaction.

The interaction between style of music and the extent to which justifications for responses involved Tempo/Rhythm/Speed or Instrumentation (Categories [i] and [ii]).

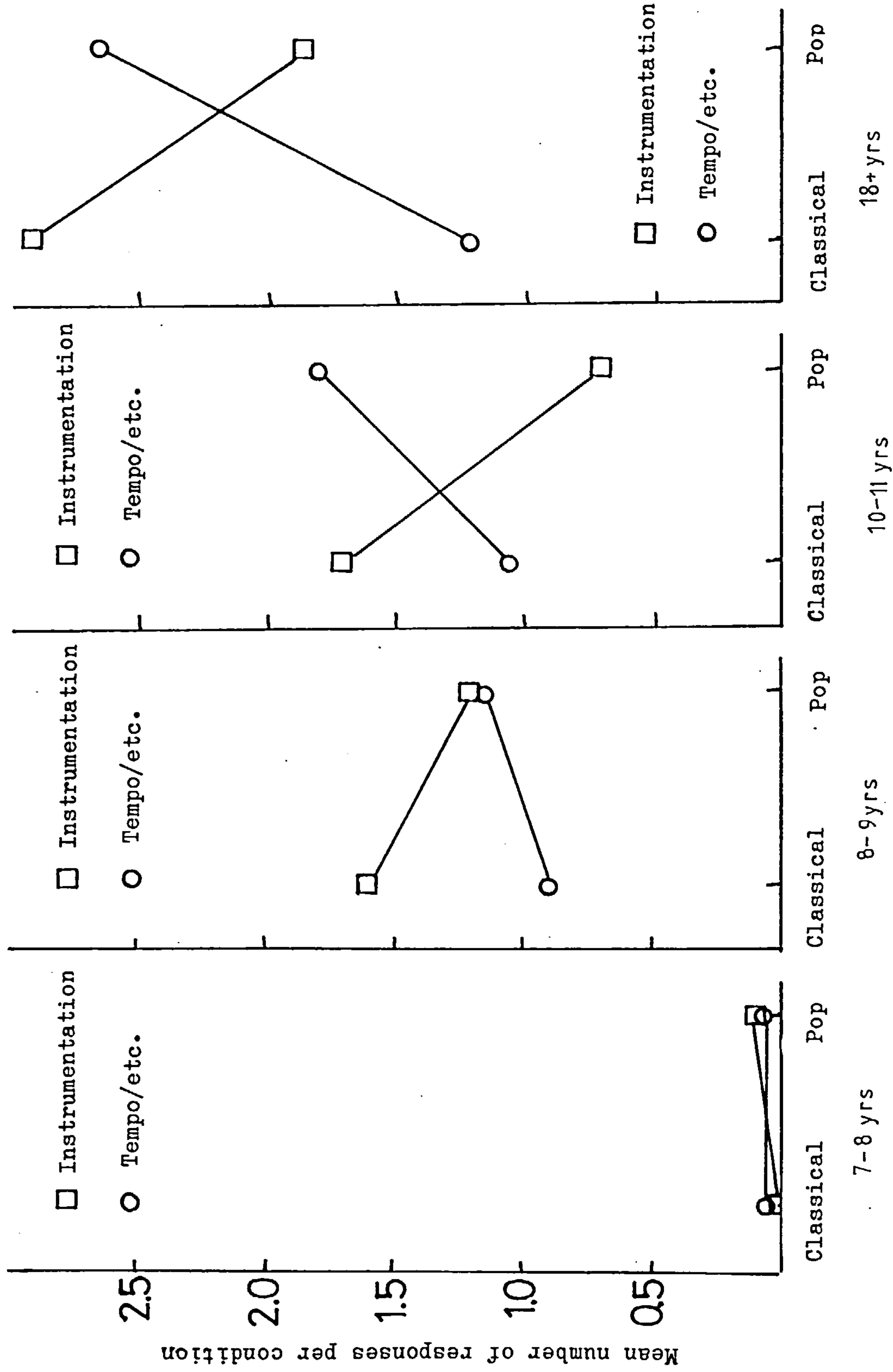


Figure 2.4.5: Significant three-way interaction. The interactions between age of subjects, style of music, and the extent to which justifications for responses involved Tempo/Rhythm/Speed and Instrumentation.

Table 2.4.1

Analysis of variance summary table, N=80  
The effects of age (4 levels), style of music, and type of stimulus pair on style sensitivity (first operationalisation).

Source of Variance	d.f.	Mean Square	F-ratio	p
A	3	4.011	10.91	<0.001
ERROR	76	0.368		
B	1	11.628	46.18	<0.001
AB	3	1.661	6.60	<0.001
ERROR	76	0.252		
C	1	0.153	0.24	n.s.
AC	3	2.536	3.94	<0.05
ERROR	76	0.645		
BC	1	0.528	1.82	n.s.
ABC	3	0.378	1.30	n.s.
ERROR	76	0.291		

A=age, 7-8 years, 8-9 years, 10-11 years, and over 18 years (independent factor).  
B=style of music, classical and popular (repeated measures factor).  
C=type of stimulus pair, "same" and "different" (repeated measures factor).

Note. Style sensitivity was operationalised as the accuracy with which subjects identified excerpts as being from the same or from different compositions.

Table 2.4.2

Frequencies with which subjects' reasons for their responses fell into categories (i) to (vii), N=80.

Style of music	Category of response	7-8 yr-olds	8-9 yr-olds	10-11 yr-olds	Over 18 yrs
Classical	(i)	2	18	23	25
	(ii)	2	32	34	59
	(iii)	1	18	12	17
	(iv)	4	10	17	28
	(v)	2	0	0	19
	(vi)	109	38	45	15
	(vii)	0	11	3	43
Popular	(i)	1	23	36	55
	(ii)	4	25	14	38
	(iii)	1	14	13	21
	(iv)	1	8	6	22
	(v)	4	0	5	19
	(vi)	109	48	47	14
	(vii)	0	7	5	40

- Key to categories:
- (i) = Tempo/Rhythm/Speed
  - (ii) = Instrumentation
  - (iii) = Timbre/Pitch/Melody
  - (iv) = Associative/Affective
  - (v) = Categorical
  - (vi) = Similarity
  - (vii) = Other

Table 2.4.3

Analysis of variance summary table,  $N=80$

The effects of age and style of music on the use of responses involving (i) Tempo/Rhythm/Speed and (ii) Instrumentation.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	3	56.111	19.37	<0.001
ERROR	76	2.898		
B	1	52.903	102.54	<0.001
AB	3	0.511	0.91	n.s.
ERROR	76	0.516		
C	1	54.853	29.72	<0.001
AC	3	6.534	1.18	n.s.
ERROR	76	1.846		
BC	1	79.298	124.83	<0.001
ABC	3	6.686	10.44	<0.001
ERROR	76	0.64		

A=age, 7-8 years, 8-9 years, 10-11 years, and over 18 years (independent factor).

B=style of music, classical and popular (repeated measures factor).

C=category of response used, Tempo/Rhythm/Speed or Instrumentation (repeated measures factor).



Table 2.4.4

Analysis of variance summary table,  $N=80$

The effects of age and style of music on the use of responses which give no information as to why two excerpts sound the same or different (Category [vi]).

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	3	171.640	42.62	<0.001
ERROR	76	4.027		
B	1	78.556	138.28	<0.001
AB	3	1.056	1.86	n.s.
ERROR	76	0.568		

A=age, 7-8 years, 8-9 years, 10-11 years, and over 18 years (independent factor).

B=style of music, classical and popular (repeated measures factor).

The direction of the main effect for style of music supported that found in the first experiment: there were more correct responses for popular than classical music. However, this effect did not hold for all age groups, as shown by the finding that there was a significant interaction between age and style. As shown in Figure 2.4.2, the adult subjects performed equally well whether excerpts were popular or classical. Thus the ability to tell whether or not two excerpts were from the same composition was affected by the style of the music until a certain age was reached, this age being somewhere between 11 and 18 years. Further research would illuminate this: all that can be said from the present findings is that the effect had disappeared by the age of 18.

Another factor which affected the performance of the four age groups differentially was the make-up of the experimental pairs (i.e. whether excerpts within a pair came from the same source or from different sources). Both adults and 7- to 8-year-olds made more Type I than Type II errors, whereas errors made by the 10- to 11-year-olds, and, to a lesser extent, the 8- to 9-year-olds, tended to be of Type II rather than Type I (Figure 2.4.3). With respect to the youngest group, Gardner (1973) remarked upon a similar finding: his 6-year-olds required excerpts to sound almost identical before they would say that they might be from the same composition, and the 7- to 8-year-olds in the present experiment had a similar tendency in that they obviously favoured the answer "no" when asked whether excerpts had the same source. It had been hoped to avoid misunderstanding of the meaning of "same"

and "different" by setting the task in a story-like context to help to make it clear even to the youngest subjects that they were not being asked to say whether or not excerpts sounded identical, and the fact that they performed above chance level suggests that this strategy was at least partially effective.

The 8- to 9-year-olds were more prepared to answer in the affirmative than the negative when asked whether or not excerpts came from the same composition, as is shown by the fact that they made more errors on "different" than on "same" pairs. This tendency was even more apparent in the case of the 10- to 11-year-olds, as can be seen in Figure 2.4.3. On the basis of the way in which this effect changed over the age groups, it could be hypothesised that the youngest subjects had relatively little awareness of the concept of style and therefore found the task difficult, although they did realise that if several excerpts were taken from one composition they were not necessarily identical. (If they were not aware of this they would have consistently answered "no" to all questions, because no two excerpts used actually were identical. This would have resulted in scores of 0 for all "same" pairs and 3 for all "different" pairs.) The 8- to 9-year-olds appeared to have greater sensitivity in that they were able to respond more accurately than the 7- to 8-year-olds to both types of stimulus pair, while the 10- to 11-year-olds showed what amounts to almost an over-awareness of the fact that passages within a composition may vary considerably: they were very ready to say that excerpts which in fact had different sources might have come from the same



source. It could be argued that having made an initial discovery of the concept of style at around the age of 9 years, the 10- to 11-year-old subjects then took an extreme viewpoint which was opposite to that held by the 7- to 8-year-olds in that their concept of style became too broad and allowed for too much variation under one style. By the time adulthood was reached responses were more accurate, although the adults in this experiment were slightly cautious in their judgements, being more ready to perceive differences in excerpts with the same source than they were to perceive similarities in excerpts with different sources. Again, it would be useful to have data from subjects between the ages of 11 and 18 to find out at what stage the high level of accuracy shown by the adult group emerges.

When an examination was made of the reasons subjects gave for their decisions, several interesting points emerged. From Table 2.4.2 it can be seen that for both classical and popular music the youngest age group appeared to find it difficult or unnecessary to explain why they gave the responses they did: many more "nonspecific" reasons, i.e. those in category (vi), occurred in this age group than in any other. There was also a difference in the frequency with which "nonspecific" reasons occurred with respect to classical and popular excerpts; as found in the first study using only two age groups, there were fewer "nonspecific" reasons for decisions about classical than about popular music. Because category (vi) included all responses which contained no information beyond the fact that excerpts "sound the same" or "sound different" this finding

suggests that subjects were less articulate about popular than classical music.

As might be predicted on the basis of the finding that 7- to 8-year-olds were likely to give "nonspecific" reasons for their responses, the number of reasons involving either tempo, rhythm, and speed or instrumentation increased dramatically with age. There were fewer reasons involving any of these variables for popular than classical pairs of excerpts; and this concurs with the hypothesis that subjects are less articulate about popular than classical music. Over both styles of music subjects were more likely to mention instrumentation (category [ii]) than tempo, rhythm, or speed (category [i]) but Figures 2.4.4 and 2.4.5 illustrate that this was not a straightforward main effect. In the case of the 7- to 8-year-olds, so few reasons were given in these categories that any difference between classical and popular excerpts in the number of reasons falling into a given category is of little importance. All remaining age groups were more likely to give reasons falling into category (ii) than category (i) where the classical excerpts were concerned: that is, they were more likely to mention instrumentation than the tempo, rhythm, or speed of a classical excerpt. The 8- to 9-year-olds, when making decisions about popular music, were equally likely to mention instrumentation or tempo, rhythm, or speed. However, the two oldest groups mentioned tempo and rhythm far more frequently than instrumentation, in the popular condition. Thus it appears that there was a tendency (which increased with age) to notice and comment upon



instrumental rather than rhythmic aspects of classical music, and, conversely, for rhythmic characteristics to be more noticeable than instrumental ones for popular music.

This finding echoes the many differences between reactions to classical and popular music which emerged during the course of this and the previous experiment. The implications of these differences are considerable and suggest many possible lines of investigation: it is unclear whether differences arise from characteristics of the music itself such as its harmonic and melodic make-up, its instrumentation, or its rhythm; or whether they might be due partly to sociological factors such as a desire to conform to peer group norms. It is also unclear whether such a distinction between musical and extramusical determinants of perceived differences is a useful one to make: this in itself needs to be established. Nevertheless the fact remains that any definition of what constitutes a musical style needs careful consideration and validation through the use of a wide range of musical material. Some of these issues will be investigated later in this thesis.

## 2.5: Conclusions (Part 2)

As already pointed out, it is apparent from the results of the first two experiments that there may be important differences in the way children respond to popular and to classical music, at least so far as these styles were represented by the compositions used here. This is illustrated quantitatively in Experiments 1 and 2 where it was found that all subjects except the adult group were more sensitive to stylistic differences (as operationalised here) when popular music rather than classical music was the source of the stimulus material being judged. Qualitative support for this notion is provided by an examination of subjects' justifications for their responses: 8- and 9-year-olds were more articulate about classical than popular music, and 10- and 11-year-olds and adults tended to comment upon different characteristics in each style of music; specifically they were more likely to refer to tempo and rhythm than to instrumentation when judging popular compositions, the reverse tendency appearing where classical stimuli were concerned. When it is also taken into account that the subjects rated the popular compositions as more familiar than the classical ones, even although there was evidence to suggest that they felt familiar with the general style rather than with the specific stimuli chosen, the implications are that future researchers should incorporate popular as well as classical music into the design of their experiments unless the generality of their

findings is to be limited.

A further general conclusion which may be drawn on the basis of the first two experiments' findings is that style sensitivity, at least as operationalised here, increases with age between 7 years and adulthood, but there were some anomalous findings in that 10- and 11-year-old subjects responded somewhat less accurately than the 8- and 9-year-old subjects with respect to popular music. It was suggested that this may have been because of their conception of the inclusiveness of one composition. A discussion of the issue in terms of categorisation might be fruitful: if it were realistic to think of distinctions between compositions or styles of music as analogous to distinctions between categories, then work on human categorisation such as that by Rosch (1978) might provide an alternative approach to the problem. This issue will be discussed in Part 4.

An examination of the distribution of Type I and Type II errors is also relevant to a discussion of the way in which style sensitivity develops. Figure 2.3.2 shows that although the ability to respond accurately to "same" pairs increases with age, the ability to respond accurately to "different" pairs actually appears to decline steadily between the ages of 7 and 11 years. Only at adulthood were subjects likely to correctly identify excerpts from different sources as being from different sources. Again, further understanding might be achieved by an examination of the way in which the ability to categorise other types of stimuli develops. One implication of the present results is that from the age of 7 years (and



possibly earlier) children's ideas about what may be considered as a member of a given aesthetic category (in this case, a composition of music) might change in such a way that they define the limits of the category in successively broader terms with increasing age, until some time between the ages of 11 and 18 years.

The preceding three paragraphs relate the results of Experiments 1 and 2 to the first two aims described in Part 2.2. The third aim was to investigate the age distribution of different criteria used by subjects to justify their responses. The most noticeable finding here was that subjects of all ages did not find it easy to express why they had responded in a given way: this was particularly true of the youngest age group, but postexperimental comments from the adult subjects suggested that they also found difficulty with this task. Where a response was given, particularly with respect to the non-adult age groups, it was likely to be in terms of a specific stimulus characteristic rather than in terms of the subjective effect the music produced on them as listeners. These findings do not entirely concur with what was hypothesised in Part 2.2: it was suggested there that older subjects might give more Analytical and fewer Associative or Affective responses than younger subjects, who might give more Associative than Analytical responses; whereas here, for all age groups, Analytical responses (categories [i], [ii], and [iii]) occurred more frequently than Associative or Affective ones (category [iv]). This is perhaps not surprising considering the nature of the task, which many subjects may

have perceived as amounting to a request to make similarity judgements between pairs of excerpts of music. The adults' responses were interesting in that they gave more responses falling into category (vii) ("Other") than did the three younger age groups. This was probably a reflection of their larger vocabularies and greater listening experience gained with age.

Overall, criteria reported to be the bases of subjects' decisions increased both in quantity and with respect to the diversity of factors involved, with increasing age.



## PART 3: AESTHETIC REACTIONS TO MUSIC:

### A COGNITIVE/EXPERIMENTAL APPROACH

#### 3.1: Literature Review

There is evidence suggesting that in a relatively unstructured task situation, subjects' written responses to musical stimuli may be of several different types. For instance, Bulloch (1921) identified four types of apperception, in research which was initially concerned with colour appreciation and which he later extended to include reactions to music. Hargreaves and Colman (1981) described five possible distinct types of response which emerged when subjects were required to state ways in which two excerpts of music were alike, and different from a third excerpt. These reactions could be roughly subdivided into those which were "objective", technical, and analytical; and those which were "subjective" and personal; the first type of response being concerned with the intrinsic qualities of the music itself such as its instrumentation, tempo, and intrinsic stylistic characteristics; the second being concerned with the effects of the music on the perceiver, which might have emotional, associative, or evaluative components. Correlation coefficients suggested that "objective" responses were more likely to be given by subjects with, than those without, some

musical experience. "Subjective" responses were more likely to arise in the musically naive than in the musically experienced.

In research into style sensitivity reported in Parts 2.3 and 2.4 of this thesis, subjects were asked to say why they thought two excerpts did or did not have the same source. No examination was made of the difference between musically experienced and musically naive subjects, but it did appear that an increase in age had an effect comparable to that of an increase in musical experience, in that adult subjects gave many more "objective" responses than 8- to 9-, and 10- to 11-year-olds, who in turn gave more of this type of response than 7- to 8-year-olds. However, it should be pointed out that the number of Associative/Affective responses also increased with age (Table 2.4.2). The reverse held for "non-specific" responses (those responses which gave no information other than that excerpts sounded as if they did or did not come from the same composition). Bartlett (1973) found that if a nondirective method of obtaining responses was used, subjects were able to discriminate between various structural elements within a composition to which they were exposed repeatedly. The common feature of the research mentioned is that subjects do, if given the opportunity, spontaneously comment upon certain intrinsic characteristics of excerpts of music to which they are exposed; and the extent to which they do this appears to be affected by both their age and their musical experience.

In the field of person perception, Honess (1981) pointed

out that it is important to use relatively unrestricting instructions in order to fully understand reasons behind subjects' affective responses to stimuli (in this case, childrens' likes and dislikes for other children). To illustrate this, he cited a study in which the authors (Livesley and Bromley, 1973, p.119) reported that a person is usually liked or disliked for his personality and behaviour rather than for his peripheral qualities such as appearance. In fact the authors specifically requested subjects to avoid using peripheral qualities in their descriptions, so this finding constituted a totally artificial effect. There is no reason to assume that it is not equally important to avoid overprescriptive instructions which may reflect experimenter preconceptions when examining reasons children give for decisions about excerpts of music.

Open-ended procedures give rise to an interesting line of inquiry. Hargreaves and Colman (1981) and Experiments 1 and 2 in this thesis find an increase with age or experience in the number of descriptive constructs used by subjects to refer to the stimulus material: this is intuitively predictable, but the question that remains, particularly in the field of music appreciation, is whether children and untrained subjects omit to mention certain characteristics of the music because they simply do not perceive them; or because, although they perceive them, they lack an adequate vocabulary with which to describe them. For instance, did the 7- and 8-year-old children in Experiment 2 tend to give no reason (or "non-specific" justifications) for their responses because



although they could hear precisely how the sounds of the instruments differed in two different excerpts, they did not know the names of instruments so they were limited to saying "it sounds different"? Was it, on the other hand, because they could not perceive precisely where in the fabric of the compositions the differences lay, but were aware that they existed at a very general level? These questions will be investigated, to differing degrees, in each of the three experiments described in Part 3.

Experiments 1 and 2 required subjects to give relatively objective judgements about excerpts of music. In everyday life it is rarely necessary to make decisions of this kind (such as deciding whether or not two excerpts come from the same composition): encounters with music are more commonly concerned with preference decisions such as choosing which record to listen to or which concert to go to. The study of the determinants of people's liking for music can be described as "aesthetics" research, in the sense of the word discussed in Part 1.2. In Part 3, three studies representing two approaches to musical aesthetics research will be described. The first of these approaches emphasises stimulus variables, such as complexity; and the second emphasises perceiver variables, such as affective state.

There is a fairly substantial body of evidence suggesting that liking for music is influenced by the complexity of the music in question. For instance, McMullen (1974) found that school-aged subjects liked melodies made up from a 5-7 pitch alphabet better than melodies having a 12 pitch alphabet, and

that a low level of redundancy was preferred to a high level of redundancy. (Melodies with increasingly high levels of redundancy were defined by McMullen as those in which the probability of the occurrence of one pitch in the alphabet, designated a dominant pitch, increases; while the probability of the occurrence of the remaining pitches in the alphabet equally decreases.) McMullen argued that mean preference scores indicated that melodies with an intermediate level of complexity were preferred over those in either the high or the low complexity categories.

McMullen and Arnold (1976) investigated the effect on preference and interest ratings of one aspect of rhythmic, rather than melodic, complexity, and found that preference tended to be an inverted U-shaped function of the amount of distributional redundancy in rhythmic sequences, and that interest responses appeared to be an increasing monotonic function of distributional redundancy. Redundancy levels were determined by using seven out of eight rhythmic sequences an equal number of times, but all seven were used less often than the eighth figure in a given composition.

Two more studies indicating that listeners are very aware of the rhythmic component of sequences and might therefore use rhythmic characteristics to a significant degree as a basis upon which to make decisions about music, are those by Heyduk (1975) and Conley (1981). Heyduk, like McMullen (1974) and McMullen and Arnold (1976), used compositions which had been specially constructed for the purpose of the experiment: in this case the stimuli were four piano compositions which



represented differing degrees of complexity as defined by their chordal and rhythmic properties. He found liking for the compositions was a unimodal function of their complexity and pointed out that this finding is consistent with predictions of an optimal complexity model of music preference. He proposed that:

Degree of preference for an event is postulated to be inversely related to the distance between the event's psychological complexity and an individual's optimal (preferred) complexity level. Thus, an inverted-U function should occur when a subject expresses preference for a range of musical events including some that are more complex and others that are less complex than optimal. If the samples from a dimension of musical complexity are more limited in range, such that the event closest in psychological complexity to the person's optimal complexity level is the most or least complex of the set, then monotonic increasing or monotonic decreasing functions relating preference and complexity are the respective expectations. (p.84).

He borrows the term "psychological complexity" from Walker (1973) who uses it as a label under which all stimulus attributes affecting aesthetic reactions to music (e.g. novelty, stimulus complexity) may be subsumed.

Unlike Heyduk (1975), McMullen (1974), and McMullen and Arnold (1976); Conley (1981) took samples of naturalistic

music as her stimulus material; this consisted of 16 of Beethoven's Eroica Variations, Op.35, for piano. Subjects were required to make judgements as to the complexity of the excerpts, and she found that these judgements correlated most highly with the rate of rhythmic activity of the music as opposed to variations in other musical parameters. Both harmonic and rhythmic variables were of importance to graduate music students in judging complexity whereas only rhythmic activity variables were of importance to less musically sophisticated subjects. She did not ask subjects to give any indication of their preferences, but her results are a convincing demonstration that subjects, especially those who have received relatively little musical training, are very aware of the rhythmic aspects of excerpts of music.

There is a relative scarcity of research into the extent to which a person's affective state might influence music appreciation. LeBlanc (1982) points out that a listener's mood can exert a subtle influence on the way he or she processes music, and that mood may also interact with stimulus characteristics. For instance, he says,

"...when an irate father tells his children to decrease the loudness of the music they are playing, the father's current affective state has interacted with a physical property of the music stimulus presented by his children." (p. 35).

He cites only two studies (Sopchak, 1955; Eagle, 1971) which

have examined the influence of current affective state on a listener's responses to music. Cantor and Zillman (1973) observed the effect on preference ratings for songs of showing to subjects film segments which varied along two dimensions: hedonic tone (i.e. films with positive or negative affective characteristics) and excitatory potential (highly arousing or not arousing). They found that a song heard immediately after presentation of a film was perceived as more pleasant, the less pleasant the film had been. The excitatory potential of the film had no significant effect on preference judgements. However, if subjects were asked to judge a song 2 minutes and 15 seconds after seeing the film no effect was observed for hedonic tone, although subjects gave significantly higher ratings in the conditions where they had been exposed to high-excitation films than in those conditions where the films had not been arousing. This research provides some evidence that perception of music may be influenced by the mood of the perceiver. However, Cantor and Zillman found that no effect on preference occurred for songs heard 4 minutes 45 seconds after the affect-inducing film, so the phenomenon was in this case short-lived.

Konecni and Sargent-Pollock (1976) and Konecni (1979) used an alternative method of inducing affect in subjects in what Davies (1982) describes as "...a series of anti-social (rather than social) psychology experiments conducted under socially bizarre circumstances in a laboratory." (p.41). Konecni (1979) argued that "...a thorough understanding of aesthetic behaviour cannot be achieved without examining how



it changes as a function of its immediate social antecedents." (p.151) and, in the experiments referred to by Davies, undertook an investigation of the effects of induced anger on preference for melodies differing in complexity. He hypothesised that preferences for the different melodies would be a function of the amount of subjects' available processing capacity. It would follow that subjects who had been angered (by being systematically insulted by a confederate of the experimenter) and hence had less available processing capacity would prefer simpler melodies because they required less processing effort than complex melodies, and this is what his results demonstrated. He offered an interpretation in terms of processing capacity rather than level of physiological arousal, because subjects who were physiologically aroused (as measured by blood pressure) but not angry did not prefer simple melodies. Konecni argued that his findings are relevant to "real-life" situations in that they suggest that musical preference may be substantially affected by the mood of the listener. He also found evidence suggesting that aggressive behaviour in angered subjects could be intensified or attenuated according to the type of music presented to them after they had been angered: those who listened to simple melodies at low volume displayed significantly less aggression than did angered subjects exposed to no stimulation; whereas exposure to complex melodies at low volume, and both simple and complex melodies at high volume, resulted in considerable amounts of aggression.

Although Konecni's research is unique in that it throws



light on a complex relationship in which the mood of the listener not only influences musical preference but may in turn be influenced by the type of music the listener is exposed to, Davies (1982) argued that his procedures are ethically inappropriate. To quote Davies, "...subjects are insulted, reviled, have loud noises blasted at them, and are then made to deliver electric shocks (or so they believe) to bystanders by way of a dependent variable." (p.41). Konecni made no mention of rhythmic complexity, which may play an important role in determining musical preference, as suggested by the research already mentioned in this review. The next experiments investigate the effects of complexity (including rhythmic complexity) upon affective and other responses to music, and Experiment 5 also includes some examination of the effects a listener's mood has on his/her responses to music, using a methodology which is, it is to be hoped, less ethically suspect than that of Konecni.

### 3.2: The Present Research

Experiments 3, 4, and 5 described in this section differ from Experiments 1 and 2 in that the stimuli used consist of experimental rather than naturalistic music. By composing stimulus sequences which varied in precise, objectively defined ways, a degree of control over the internal stimulus characteristics can be attained which means that any changes in subjects' responses can be directly attributed to specific changes in the musical material. It was not possible to do this with respect to responses given by subjects in Experiments 1 and 2. Experiments 3 and 4 both make use of the same stimulus sequences, which involve the manipulation of three musical parameters, namely rhythmic complexity, melodic complexity, and tempo. In Experiment 5, rhythmic complexity alone is manipulated. There follows a summary of the aims of each experiment.

The primary aim of Experiment 3 was to investigate further the issues discussed in Part 3.1 relating to the development of the criteria upon which similarity/difference judgements are based, in the field of music perception. Instead of subjects being required to describe in writing what it was about the music that made them decide whether two excerpts came from similar or different sources, the stimulus material was designed so that changes in any of its physical characteristics which might have influenced such a decision were precisely defined from the outset. The degree to which

these characteristics actually did influence similarity judgements was ascertained by requiring subjects to give a simple numerical rating of how similar or different a given pair of sequences appeared to be, in the context of a role-playing situation which it was hoped would clarify the nature of the task for the younger subjects. Ratings were given nonverbally in order to avoid any apparent differences that might emerge in perceptual ability because of actual differential verbal abilities among the four age groups tested (7- and 8-year-olds, 10- and 11-year-olds, 13- and 14-year-olds, and adults). Only secondarily were subjects invited to make some verbal comment, this being entirely voluntary.

The results of Experiments 1 and 2 (present thesis) and those of Hargreaves and Colman's (1981) study suggest that subjects are more likely to justify their decisions about the perceived similarity of excerpts by giving analytical statements than they are to do so by discussing the affective qualities of the music or what categories it belongs to. In other words they are more likely to say of two compositions "the tune is different" than they are to say "one is medieval and the other isn't" or "one is happy and the other sad". For this reason, no overt attempt was made to alter the affective or stylistic qualities of the stimulus sequences used. The parameters selected (melodic complexity, rhythmic complexity, and tempo) were chosen for two reasons. Firstly, it was found in Experiments 1 and 2 that subjects of all ages often mentioned melody, rhythm, or tempo as criteria for their



responses, and it was therefore assumed that they were salient features of the music in this respect. Secondly, all of these stimulus characteristics have the property of being variable along a continuum (i.e. less to more complex, or slower to faster). For this reason, although instrumentation was often mentioned in Experiments 1 and 2 as a criterion, it was not one of the parameters included: the differences between (for instance) a guitar sound and the sound of a flute are discrete rather than continuous, and cannot easily be varied along one dimension.

Seven experimental stimulus sequences were produced. One of these was a standard, in which melodic complexity, rhythmic complexity, and tempo were all at a moderate level (according to objective measures described in Appendix 3.3.1). In three of the other sequences, either melodic complexity or rhythmic complexity or tempo were increased; and in each of the remaining three sequences one of these parameters was decreased. Increments and decrements in any one parameter were equal in magnitude although opposite in direction when compared with the standard. Only one parameter was altered at a time: for instance, in the variation which was more complex melodically than the standard sequence, rhythmic complexity and tempo were the same as they were in the standard sequence.

Each of the six sequences in which a parameter had been varied were paired one at a time with the standard sequence, and subjects from four age groups rated the extent to which members of a pair sounded similar or different. The hypotheses under investigation were: (1) whether (a) an increase in age



or (b) the presence or absence of musical training would be associated with an increase or decrease in sensitivity to differences between sequences, using a largely nonverbal measure of sensitivity; and (2) whether an increase in rhythmic complexity, melodic complexity, or tempo was perceived as being equal to a decrease (equal in magnitude, objectively, to the increase) in the same parameter. The interaction of these factors was also examined.

In Experiment 4, the same seven stimulus sequences were used. However, the emphasis in this case was upon subjective impressions created by the sequences rather than upon the relatively objective judgement of similarity between sequences. The aims of the experiment were twofold: an examination was made of the extent to which the objective measures of complexity used to generate the sequences were reflected in subjects' actual perceptions of complexity. The second aim was to investigate subjects' liking for each sequence, relative to that for the standard. In both cases responses given by musically experienced subjects were compared with those given by inexperienced subjects, because it was hypothesised on the basis of Conley's (1981) study that judgements of complexity might be influenced by extent of musical training and experience, and it did not seem unreasonable to suppose that liking for the sequences might be similarly affected, especially in the light of Heyduk's (1975) discussion of individuals' optimal complexity levels.

In Experiment 5, rhythmic complexity was the only stimulus variable which was manipulated. In each of

Experiments 1 to 4, the emphasis has been on the extent to which reactions to music, whether naturalistic (Experiments 1 and 2) or experimental (Experiments 3 and 4), are influenced by variations in intrinsic features of the music aside from extrinsic, environmental features. In Experiment 5 perceiver characteristics were taken into account in addition to stimulus characteristics, in that an investigation was undertaken into the effects on subjects' preferences of imagining that they were feeling angry while listening to the stimulus sequences. To some extent the study followed on from those of Konecni and Sargent-Pollock (1976) and Konecni (1979), but it differed in the two respects already mentioned: namely, rhythmic complexity rather than melodic complexity was manipulated, and instead of actually being made angry, subjects were simply asked to imagine that they were angry. Although problematical, this second approach was adopted as an alternative to that used by Cantor and Zillman (1973) to induce an affective state in their subjects, because their technique (i.e. that of showing films with a given hedonic predetermined hedonic and excitatory effects) produced only transitory effects which may not necessarily have been true to life.

One further measure was taken to give the experimental situation more resemblance to situations in which music is normally heard. McMullen and Arnold (1976) did not give any details of a melodic component to their rhythmic sequences, and because it is uncommon in Western music to find a purely rhythmic composition with no melody, it was felt that the

sequences in the present experiment would be more like natural music if they consisted of both a melody and a rhythm. They were composed with this in mind, the complexity of the former being held constant so as not to confound any effect of rhythmic complexity.

To summarise, the null hypotheses tested in Experiment 5 were as follows:

1. Subjects imagining that they are angry, and a control group who merely listen to the sequences, will display equal distributions of preferences for rhythms of low, intermediate, and high complexity, and silence.

2(a). Within the control group (i.e. those subjects not imagining anger), preferences for rhythms of low, intermediate and high complexity, and silence, will be evenly distributed, for subjects who describe themselves as "very calm" at the time of the experiment. (b) The same will apply for those subjects who describe themselves as "slightly irritated" or "very irritated".

3(a). Within the experimental group (i.e. those subjects imagining anger) preferences for rhythms of low, intermediate and high complexity, and silence, will be evenly distributed for subjects who describe themselves as "very calm" at the time of the experiment. (b) The same will apply for those subjects who describe themselves as "slightly irritated" or "very irritated".

4. Within the control group, preferences for rhythms of low, intermediate, and high complexity, and silence, will be evenly distributed, for each of the age groups tested.



5. Within the experimental group, preferences for rhythms of low, intermediate, and high complexity, and silence, will be evenly distributed for each of the age groups tested.

There were in addition two more hypotheses which investigated the extent to which preferences actually were dependent upon rhythmic complexity rather than melodic complexity or any other variable. These were as follows:

6. Preferences for melodies, regardless of their rhythmic accompaniment, will be evenly distributed.

7. Reasons given by subjects for their preferences are equally likely to involve melody, rhythm, emotional/associative connotations or other aspects of the music.



### 3.3: Experiment 3

#### The Effects of Variations in Complexity on the Perceived Similarity of Quasi-Musical Sequences

#### Method

##### Subjects.

There were 80 subjects divided into four groups of 20 by age, the youngest being 7 to 8 years old, the second 10 to 11 years, the third 13 to 14 years, and the fourth 18 years and over. In each age group there were approximately equal sex ratios, there being 47 females and 33 males altogether. The three younger groups were taken from schools in Leicestershire and tested in groups of between five and ten members. The adults were volunteers, many being students at Leicester University, and were tested in groups of two to five members, depending on availability. The mean age of the adult group was 23.7 years.

##### Musical material and questionnaires.

A sequence of music was composed especially for the experiment, using a polyphonic synthesizer and a programmable rhythm generator for the melodic and percussive components of

the sequence respectively. This sequence was to be used as a standard, and because it was to be varied systematically along three parameters (rhythmic complexity, melodic complexity, and tempo) it was necessary to obtain precise measures of each of these parameters. Melodic complexity was objectively determined by using Davies'(1969) statistical approximations to Western music (5 notes' restraint) as the source of the melodic component of the standard sequence. This was 12 bars long. A percussive accompaniment to the melody was written using the rhythm generator, and this was of moderate complexity in terms of the number of events per bar and their positions in each bar: the overall effect was of a steady, regular drumbeat with little syncopation but enough variation to retain a listener's interest. Full details are given as to the melody of and percussive accompaniment to the standard sequence in Appendices 3.3.1 and 3.3.2 respectively.

The standard sequence was recorded on a reel-to-reel tape which was used as a master. Variations of the standard sequence were then composed, each of which retained all of its characteristics except one: either melodic complexity or rhythmic complexity or tempo were changed one at a time. Since each of these parameters could be varied in either one of two directions (either more or less complex, or faster or slower), six variations resulted.

Taking each parameter in turn, precautions were taken to ensure that it was altered to the same degree in both directions; for instance, tempo was doubled relative to the standard for the first variation and halved relative to the

standard for the second variation. In the third variation the number of rhythmic events per bar was approximately doubled as was the number of events which were positioned "off the beat", giving the percussive accompaniment to the variation an irregular, disjointed character, although the melody and tempo remained the same as those in the standard sequence. The fourth variation had a very simple rhythmic accompaniment: there were half as many events per bar as there were in the standard sequence and half as many fell on "off-beats" as did in the standard. Again tempo and melodic complexity remained unchanged.

The fifth and sixth variations retained the tempo and rhythmic accompaniment of the standard sequence, but the complexity of the melody was manipulated by introducing four notes more or less restraint than applied in the case of the standard. Thus the fifth variation, which sounded complex subjectively, had a melody based on a sequence generated using only one note's restraint, while the sixth and final variation, sounding subjectively simple, had a melody taken from a sequence generated with nine notes' restraint.

The six variations were recorded one at a time on the master tape. Full details of the variations in rhythmic and melodic complexity are given in Appendices 3.3.3 to 3.3.6 inclusive.

Two experimental cassette tapes were made up. Each of these consisted of the standard sequence followed by the six variations, interspersed at some point with a repetition of the standard, with a 5-second pause between each sequence of



music. Two different random orders were used for the arrangement of the sequences. They then appeared a second time in the same order as they had been on the first half of each tape, but each one was presented in conjunction with the standard so that a direct comparison could be made of the standard with each variation and with itself. There were 5-second pauses between sequences to be compared. The final make-up of the first experimental tape was: standard, variation 4, variation 2, variation 6, variation 5, standard, variation 1, variation 3; there being a 5-second pause between each of these sequences. The ordering of sequence pairs in the second half of the tape was: standard, variation 4; standard, variation 2; standard, variation 6; standard, variation 5; standard, standard; standard, variation 1; standard, variation 3; there being a 5-second pause between the members of each "standard, variation" pair.

On the second experimental tape the order in which the sequences were presented was: variation 5, variation 4, standard, variation 1, variation 3, variation 6, variation 2. The total length of each tape was 8.5 minutes.

Subjects were required to say how similar (on first impression) each variation was to the standard. The context in which the task was set was that of a television producer's office, as described in detail in the "procedure" section. Booklets were provided in which subjects recorded their responses: these consisted of a plain front cover and seven pages, each with the question "How many marks does this 'spare' get for being a good match?" followed by a range of



ten marks from 1/10 to 10/10 of which the subject could circle one. The final question on each page was, "To help the producer, can you suggest any ways of changing the 'spare' to make it a better match?" A sample page is presented in Appendix 3.3.7.

#### Procedure.

The youngest and oldest subjects were tested in smaller groups than the 10- to 11- and 13- to 14-year-olds (3 to 5 members as opposed to 8 to 10 members); the 7- and 8-year-olds because it was easier to ensure that they had understood the task when there were few of them together, and the adults because they were volunteers and were therefore rarely available in large numbers at the same time. Subjects of school age were tested in quiet classrooms during school hours, adult subjects were tested in a soundproof room in the Psychology Department of Leicester University.

Once a group had assembled, subjects were given the following instructions by the experimenter:

For the next 20 minutes or so I would like you to imagine that you are assistants to a television producer. The job of this producer is to choose background music to various programmes: do you think that it would matter what kind of music he chose for each programme?

(Some discussion followed of how different kinds of music can have different "feelings" or evoke different atmospheres,

pointing out that, for instance, it would probably be quite inappropriate to have "Dr. Who" music playing behind a programme such as "Coronation Street". This discussion was extended in the case of the younger subjects in order to ensure that they had understood.) The experimenter then continued:

One day you come into the office to find the producer in a terrible state. He has to put out a programme that very afternoon, and he had thought that he had everything ready, but has discovered that the tape of music for the programme has been lost, and he can't find it anywhere. He searches through all his drawers and cupboards, and there's no sign of it, but he does find seven other tapes. "Aha!" he thinks, "One of these might do the job!" So he listens to the seven pieces. However, they all seem fairly similar to the lost one and he can't really decide whether any of them match it better than any of the others; so he decides to ask his assistants (you) to listen to all of them and give each one marks out of ten according to how well they think it would fit in the same part of the same programme. I'm not going to tell you what the programme is: it could be almost anything: but I will play you the lost tape because although the real assistants must know it well, you will never have heard it before. First of all, you'll hear the lost tape, then you'll hear the seven "spare" tapes which the producer found, played one after the other. There's no need for

you to decide now which are good matches and which are not, because I'll play all of them again later, one by one, each one with the lost one. So this time just listen, to get some idea of what the lost one and the spares are like. The producer just wants everybody's own ideas, so it doesn't matter if you don't think the same as your friends.

Wording varied slightly to accommodate the younger subjects and to preserve an atmosphere of informality. The cassette player was started and subjects heard the standard sequence (i.e. the "lost" tape) followed by the six variations with the repeat of the standard among them (i.e. the seven "spare" tapes). When all eight sequences had been heard, the cassette player was turned off. Subjects were given one booklet each and were asked to open them at the first page. It was explained that there was one page for each "spare" tape, and that they were to decide how good a match each "spare" tape was to the lost tape and give it marks out of 10 accordingly, scoring 10/10 if the "spare" was a perfect match, 1/10 if it was quite out of the question that the "spare" would be appropriate in the same part of a programme as the lost tape, and 5/10 if it was neither a particularly good nor a particularly bad match. The marking scheme was made more explicit if required in the case of the youngest subjects. In all cases it was stressed that the same score could be used more than once during the testing session. Subjects were asked to write a response to the second question on each page if



they had any ideas about ways in which the "spare" tape might be altered to make it more like the lost one. They were told that they would hear the lost tape, then the first "spare", and there would then be a pause while they circled the number of marks they thought the first "spare" deserved. They would then hear the lost tape again, followed by the second "spare" with another pause during which they rated the second "spare" on the second page, and this procedure would continue until they had rated all seven "spare" tapes.

After the instructions had been given, and subjects' queries had been answered, the tape was started at the point at which it had been stopped previously. After playing the standard and whichever variation appeared first on the tape (variation four on tape one, variation five on tape two), the tape was stopped again while subjects made their responses. When they had finished the questions on the first page they were asked to turn over and the tape recorder was started again, playing the standard sequence once more, followed by the next variation (variation two on tape one, variation four on tape two). The tape was stopped while subjects completed the second page of their response booklets. This procedure was repeated until all six variations, and the standard itself (disguised as a "spare" tape) had been rated according to how well they matched the standard sequence. At this point, subjects were asked to turn to the covers of their response booklets and write on them their age, sex and whether they played any instruments, sang, or were active in any other musical capacity.



Out of each age group, half the subjects heard tape one and the other half heard tape two, there being approximately equal sex ratios in each subgroup. After each experimental session, response booklets were marked according to the tape heard in order to avoid confusion when analysing responses.

### Analysis of Responses and Results

In order to discover whether subjects had recognised the standard sequence (when presented as a "spare" tape) as being a good match with itself (when presented as the "lost" tape), the mean similarity ratings were calculated for the fifth "spare" on tape one, and the third "spare" on tape two. The mean scores were 9.38 for the standard on tape one and 9.48 for the standard on tape two.

There was a possibility that a particular random order of presentation of the variations might be associated with a particular pattern of similarity ratings. In order to assess the extent to which this had happened, a Pearson's product moment correlation coefficient was calculated on the mean scores given by each of the four age groups to each of the "spare" sequences (i.e. six variations and the standard sequence), to find out whether responses to stimuli on tape one correlated with responses to stimuli on tape two, regardless of the fact that they were presented in different orders. This yielded 28 mean similarity ratings for each tape, as shown in Table 3.3.1. The correlation coefficient was 0.79,  $p < 0.001$ , 2-tailed test. Because this value was statistically

significant it was assumed that the position in which a "spare" sequence appeared on the tape bore no systematic relation to the magnitude of the similarity rating it received, so data from the two tapes were henceforth pooled.

Ratings which subjects gave to the three types of variation (of rhythm, melody, and tempo) were examined separately. Three, two-way analyses of variance were carried out (age x tempo, age x rhythmic complexity, and age x melodic complexity), with repeated measures on the second factor in each case. The summary tables for these analyses are shown in Tables 3.3.2, 3.3.3, and 3.3.4 for tempo, rhythmic complexity, and melodic complexity respectively. For the first analysis (age x tempo) there was one significant main effect: that for age ( $F[3,76] = 3.51$ ,  $p < 0.05$ ). This is illustrated in Figure 3.3.1. Neither the main effect for tempo, nor the interaction effect, were statistically significant. For the second analysis (age x rhythmic complexity) both main effects and the interaction effect reached statistical significance. For age,  $F(3,76) = 4.47$ ,  $p < 0.01$ . For rhythmic complexity,  $F(1,76) = 75.66$ ,  $p < 0.001$ . For the interaction effect,  $F(3,76) = 11.02$ ,  $p < 0.001$ . The main effect for age and the interaction effect are illustrated in Figures 3.3.2 and 3.3.3 respectively. So far as the main effect for rhythmic complexity was concerned, examination of the means showed that the variation which was more complex than the standard was seen as a better match to the standard than was the less complex variation (mean ratings were 8.24 and 5.80 respectively). For the third analysis (age x melodic complexity) there was only one significant main

effect, namely that for melodic complexity ( $F[1,76] = 12.26$ ,  $p < 0.001$ ). Examination of the means showed that the more complex variation was perceived as a better match to the standard than was the less complex variation (the mean ratings were 4.75 and 3.95 respectively).

Data from all those subjects who indicated that they had received some kind of musical training were separated out: 29 subjects out of all age groups fell into this category. From those remaining with no musical training 29 further subjects were selected, who matched the first group with respect to age and sex. Using these two sets of data, three further two-way analyses of variance were performed examining the interactive effects on similarity ratings of the presence or absence of any musical training and a change in complexity or speed of the stimulus sequences. The summary tables for these analyses are shown in Tables 3.3.5, 3.3.6, and 3.3.7 (showing effects of changes in tempo, rhythmic complexity, and melodic complexity respectively). For the first of these analyses (training x tempo), none of the main effects or interaction effects reached statistical significance. For the second analysis (training x rhythmic complexity), the main effect for rhythmic complexity was statistically significant ( $F[1,56] = 39.59$ ,  $p < 0.001$ ). Examination of the means showed that the more complex variation was perceived as a better match to the standard than was the less complex variation (the mean ratings were 8.35 and 5.90 respectively). For the third analysis (training x melodic complexity) there was only one statistically significant effect, namely that for training



( $F[1,56] = 7.78$ ,  $p < 0.01$ ). Examination of the means showed that the trained subjects perceived both melodic variations as a worse match to the standard than did the untrained subjects. The mean rating given by trained subjects was 3.70, and the mean rating given by the untrained subjects was 5.12.

Subjects' responses to the question, "To help the producer, can you suggest any ways of changing the 'spare' to make it a better match?" were categorised according to whether or not they had correctly identified the parameter which had been changed in each variation, as compared with the standard sequence. Responses to the rhythmic, tempo and melodic variations were taken separately. In each case, subjects could either answer appropriately (e.g. for variation one, which was twice as fast as the standard, an appropriate answer might be "Slow it down"), or inappropriately (for variation one there would be two possible types of inappropriate response, one involving the suggestion that the pitches of the notes should be altered, the other suggesting that the amount of rhythmic activity per bar in the percussive background should be altered). Alternatively subjects might make no response, or discuss the tape in such a way that the response did not fit into either the "appropriate" or "inappropriate" categories. An example of the latter type of response was, "Employ a new producer!". Eleven categories were used altogether, three for appropriate mention of rhythm, tempo or melody, and six for "inappropriate" responses: for each of the three types of variation under examination, there were two possible inappropriate responses arising from mention of either of the



other two types of parameter which had in fact remained unchanged. The remaining two categories comprised "other" responses and instances where there was no response.

The response, or absence of a response, to each variation by each subject was placed in one of the eleven categories. Table 3.3.8 shows the percentage of responses from each age group which fell into each category.

Theoretically it would have been possible for subjects to make a further type of inappropriate response, concerning the direction in which a parameter was varied: for instance, in the case of a variation in which the tempo had been increased they might suggest that it could be made to match the standard by increasing the tempo rather than decreasing it. However, in practice, this type of error never occurred: if subjects could correctly identify the parameter requiring modification they were never mistaken about the direction of that modification.

### Discussion

The task which subjects were required to do amounted to giving ratings of the extent to which each variation sounded similar, subjectively, to the standard sequence. The standard sequence was itself included among the "spares" as a partial check that the task was correctly interpreted by subjects: because the standard was identical to itself it was assumed that it would achieve a high score when presented in the guise of a "spare" tape to be compared with the "lost" one. Mean scores allotted to it were in fact sufficiently high to

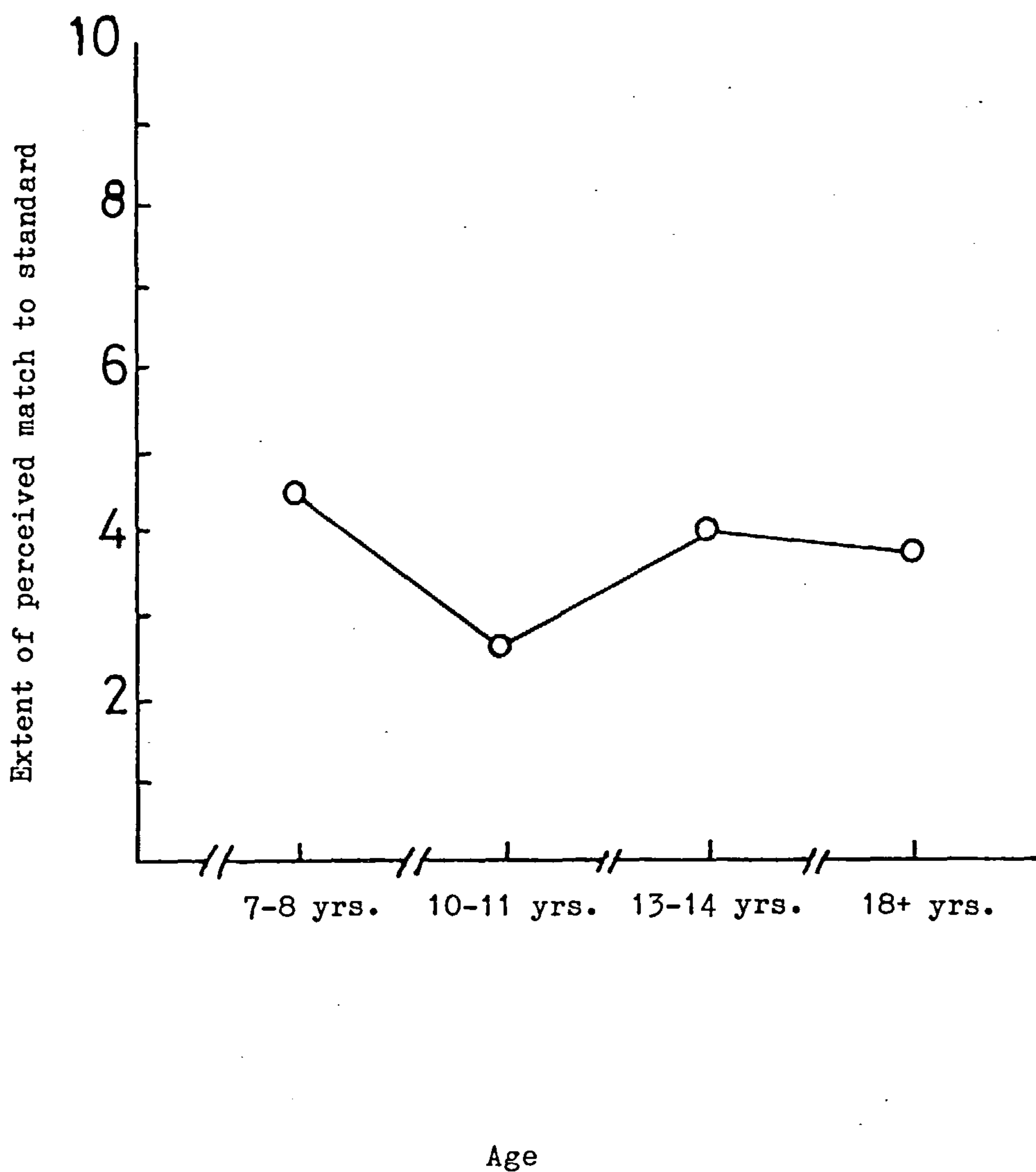


Figure 3.3.1: Significant main effect.  
The effect of age on ratings of the extent to which variations one and two matched the standard sequence. (The higher the rating, the better the perceived match).

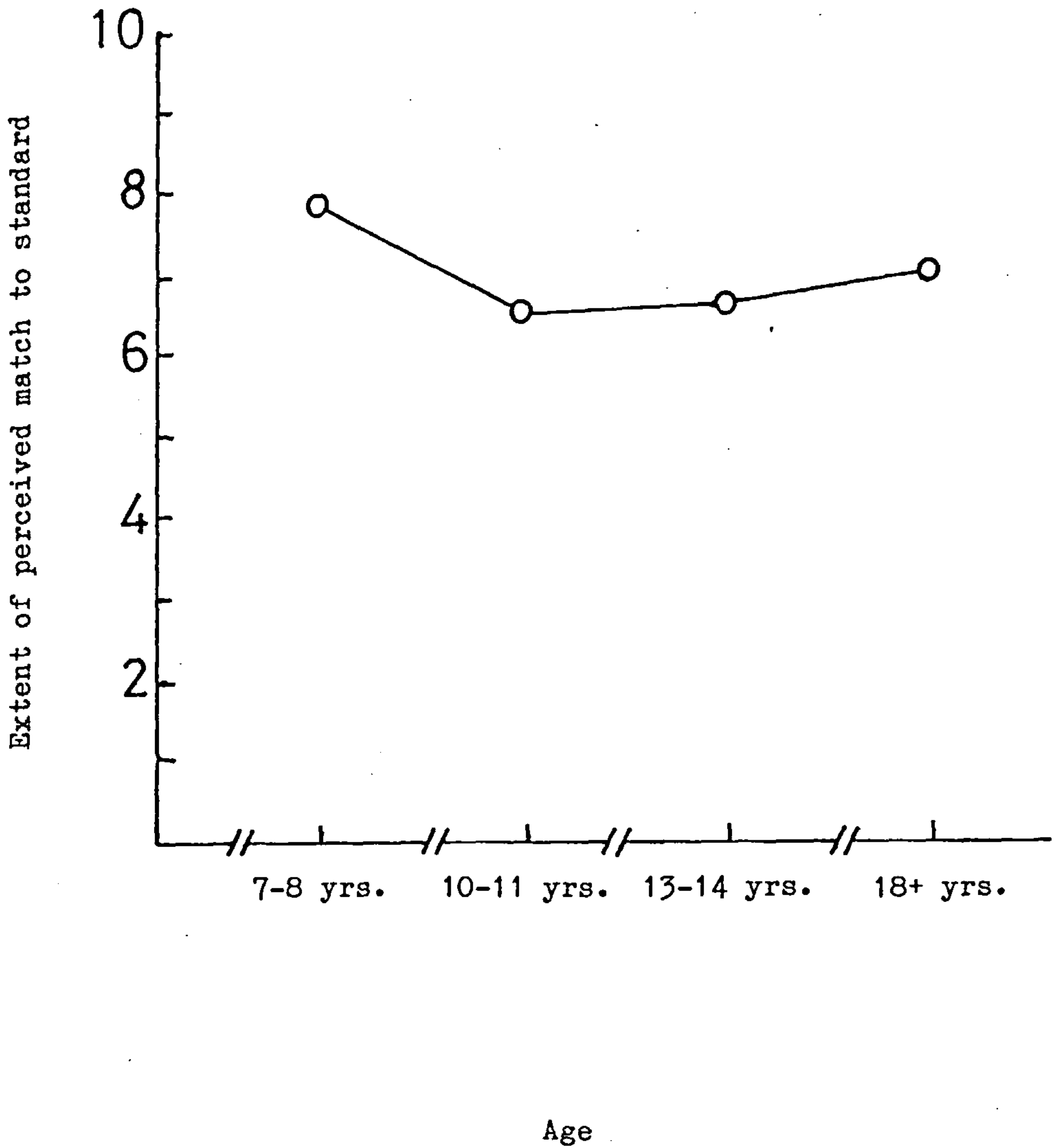


Figure 3.3.2: Significant main effect.

The effect of age on ratings of the extent to which variations three and four matched the standard sequence. (The higher the rating, the better the perceived match).

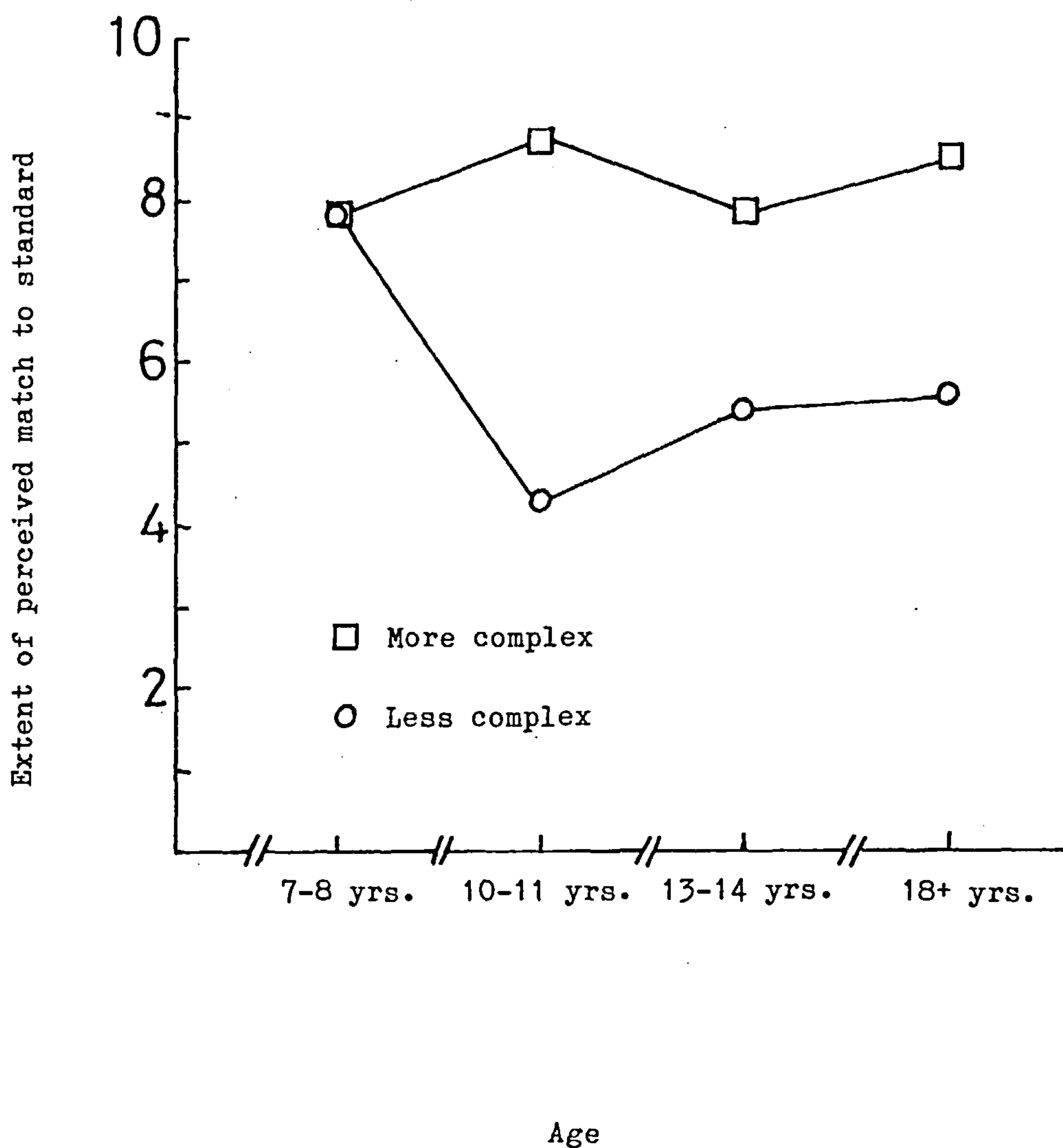


Figure 3.3.3: Significant two-way interaction. The effects of age, and whether complexity was increased or decreased, on ratings of the extent to which variations three and four matched the standard sequence.



Table 3.3.1

Mean ratings of extent to which six "spare" sequences match the standard sequence, for tape one and tape two.

Age	Type of "spare"	Tape one	Tape two
7 to 8 years <u>n=20</u>	s	9.6	9.2
	>t	4.8	3.4
	<t	6.3	3.5
	>r	8.7	7.0
	<r	8.8	7.0
	>m	4.7	4.7
	<m	4.5	2.9
10 to 11 years <u>n=20</u>	s	9.2	9.5
	>t	2.6	3.8
	<t	2.4	1.8
	>r	8.9	8.6
	<r	5.0	3.6
	>m	3.8	5.5
	<m	5.7	2.5
13 to 14 years <u>n=20</u>	s	9.9	9.4
	>t	4.0	4.0
	<t	4.0	4.0
	>r	9.0	6.7
	<r	5.7	5.1
	>m	2.0	6.4
	<m	2.8	4.4
18 years and over <u>n=20</u>	s	8.8	9.7
	>t	3.5	3.5
	<t	4.0	3.8
	>r	8.4	8.6
	<r	6.0	5.2
	>m	5.2	5.7
	<m	5.5	3.3

Key to "spares":

s=standard sequence (presented as a "spare")

>t=variation one (faster than standard)

<t=variation two (slower than standard)

>r=variation three (more complex rhythmically than standard)

<r=variation four (less complex rhythmically than standard)

>m=variation five (more complex melodically than standard)

<m=variation six (less complex melodically than standard)

Table 3.3.2

Analysis of variance summary table,  $N=80$

The effect of age on ratings of the extent to which variations one and two matched the standard sequence.

Source of Variance	d.f.	Mean Square	F-ratio	<u>p</u>
A	3	24.423	3.51	<0.05
ERROR	76	6.955		
B	1	0.156	0.05	n.s.
AB	3	6.823	2.11	n.s.
ERROR	76	3.237		

A=age, 7-8 years, 10-11 years, 13-14 years, and over 18 years (independent factor).

B=tempo, slower or faster than standard (repeated measures factor).

Table 3.3.3

Analysis of variance summary table,  $N=80$

The effect of age on ratings of the extent to which variations three and four matched the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	3	15.106	4.47	<0.01
ERROR	76	3.383		
B	1	239.256	75.66	<0.001
AB	3	34.84	11.02	<0.001
ERROR	76	3.162		

A=age, 7-8 years, 10-11 years, 13-14 years, and over 18 years (independent factor).

B=rhythmic complexity, more or less complex than standard (repeated measures factor).

Table 3.3.4

Analysis of variance summary table,  $N=80$

The effect of age on ratings of the extent to which variations five and six matched the standard sequence.

Source of Variance	d.f.	Mean Square	F-ratio	p
A	3	7.417	0.99	n.s.
ERROR	76	7.515		
B	1	51.200	12.26	<0.001
AB	3	0.683	0.16	n.s.
ERROR	76	4.176		

A=age, 7-8 years, 10-11 years, 13-14 years, and over 18 years (independent factor).

B=melodic complexity, more or less complex than standard (repeated measures factor).



Table 3.3.5

Analysis of variance summary table, n=58

The effect of musical experience on ratings of the extent to which variations one and two matched the standard sequence.

Source of Variance	d.f.	Mean Square	F-ratio	<u>p</u>
A	1	19.862	2.22	n.s.
ERROR	56	8.946		
B	1	0.034	0.01	n.s.
AB	1	0.552	0.16	n.s.
ERROR	56	3.436		

A=musical experience, extensive or minimal (independent factor).

B=tempo, faster or slower than standard (repeated measures factor).

Table 3.3.6

Analysis of variance summary table, n=58

The effect of musical experience on ratings of the extent to which variations three and four matched the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	13.112	2.89	n.s.
ERROR	56	4.534		
B	1	171.388	39.59	<0.001
AB	1	0.698	0.16	n.s.
ERROR	56	4.329		

A=musical experience, extensive or minimal (independent factor).

B=rhythmic complexity, more or less complex than standard (repeated measures factor).

Table 3.3.7

Analysis of variance summary table, n=58

The effect of musical experience on ratings of the extent to which variations five and six matched the standard sequence.

Source of Variance	d.f.	Mean Square	F-ratio	<u>p</u>
A	1	57.966	7.78	<0.01
ERROR	56	7.450		
B	1	9.966	2.62	n.s.
AB	1	0.138	0.04	n.s.
ERROR	56	3.802		

A=musical experience, extensive or minimal (independent factor).

B=melodic complexity, more or less complex than standard (repeated measures factor).

Table 3.3.8

Percentages of appropriate, inappropriate, and other responses made by each of four age groups (n=20 for each age group).

Response made	Parameter which should have been named	Percentage of responses at each age			
		7-8 yrs	10-11yrs	13-14yrs	18+ yrs
Change tempo	Tempo*	19.0*	23.5*	20.5*	21.5*
	Rhythm	9.0	14.0	13.5	12.0
	Melody	1.0	3.5	2.0	0.0
Change rhythm	Tempo	0.0	0.0	0.0	0.0
	Rhythm*	0.5*	2.0*	3.0*	10.5*
	Melody	0.0	0.0	1.0	1.0
Change melody	Tempo	0.0	1.0	1.0	0.5
	Rhythm	0.0	1.0	0.0	1.5
	Melody*	0.0*	5.5*	8.5*	11.0*
No response	Not analysed	64.0	44.5	43.5	24.5
Other response	Not analysed	6.5	5.0	7.0	17.5

Asterisks mark each row of "appropriate" responses.



justify the supposition that subjects had interpreted the instructions in the manner intended.

The effects on similarity ratings of systematically varying three musical parameters were examined. Any direct comparisons between parameters (for example, a comparison of the effect of increasing melodic complexity with the effect of increasing rhythmic complexity) were felt to be inappropriate, because there was no objective method of ascertaining whether the extent to which one parameter had been altered was equivalent in scale to the extent to which any other parameter had been altered. However, steps had been taken to ensure that for any one parameter, the incremental alteration (i.e. increase in tempo or complexity) was objectively equal to the decremental alteration (i.e. decrease in tempo or complexity), and the question of interest was therefore whether any differences existed in the subjective experience of these two types of alteration (within, as opposed to between, parameters) as measured by rated match to the standard sequence.

In the case of rhythmic and melodic complexity, the answer to this question appeared to be in the affirmative: as reported, in both cases the more complex variation achieved higher similarity ratings than the less complex variation. This finding held true for all age groups except the 7- to 8-year-olds who found that the variation which was rhythmically less complex matched the standard just as well as the variation which was rhythmically more complex than the standard (Figure 3.3.3). However, there was no such age effect

for tempo; variations in which tempo had been altered in either direction were judged to be relatively dissimilar to the standard by all four age groups.

The phenomenon whereby the more complex variations appeared to be more similar to the standard than the less complex variations, despite the fact that both variations objectively differed an equal amount from the standard, might be explained in terms of a "ceiling" effect. It is possible that the standard was itself fairly complex, subjectively, and any increase in complexity (rhythmic or melodic) over and above this was not so apparent as a decrease in complexity. This possibility will be discussed in Experiment 4 with reference to an independent measure of the perceived complexity of each variation.

Ratings for variations involving changes in tempo or in rhythmic complexity were somewhat affected by the age of the subjects. From Figure 3.3.2 it can be seen that the 7- to 8-year-olds appeared to notice changes in rhythmic complexity less than the older subjects did, because they gave both types of rhythmic variation higher similarity ratings than did the other age groups. So far as changing the tempo was concerned, this was noticed most by 10- to 11-year-olds, who found any variation in which tempo had been altered very dissimilar from the standard (mean similarity rating was only 2.65) as can be seen from Figure 3.3.1. Overall, however, age had relatively little effect on the ability of subjects to detect a mismatch between a variation and the standard sequence, especially where melodic complexity was concerned.

Previous research (Experiments 1 and 2 in this thesis, Gardner, 1973) has demonstrated a marked difference among age groups in subjects' ability to express verbally what they perceive when listening to music. The findings of the present experiment support the contention that this is not necessarily a reflection of discrepancies in perceptual ability per se: if it were, a more definite developmental pattern would have been expected to emerge in the distribution of similarity ratings for each variation as compared with the standard.

Although melodic complexity was the one parameter within which there was no main effect for age, it was the only parameter for which the presence or absence of musical training appeared to have any relevance. The musically trained subjects perceived any variation in which melodic complexity had been altered as being less like the standard sequence than did untrained subjects when judging the same variation. Thus it might be argued that musically trained subjects were aware of changes in melodic complexity to a greater extent than were untrained subjects. This is interesting in the light of work done by Conley (1981) in which she found that harmonic and melodic variables were of importance to graduate music students in judging the complexity of 16 of Beethoven's Eroica Variations, but that they were of no importance to less musically sophisticated subjects. However, all groups in Conley's study found rhythmic variables to be important criteria in making complexity judgements. The results of the present experiment do not contradict this, because whether or not subjects were musically trained had no significant effect



on the similarity ratings they gave for variations in which rhythmic complexity had been altered. This was also the case for variations in which tempo had been altered.

In summary, results of the analyses of variance showed that, over all 80 subjects, age had relatively little effect on similarity ratings except for the tendency of 7- to 8-year-old subjects not to distinguish as clearly as older subjects between an increase and a decrease in rhythmic complexity. Secondly, 10- to 11-year-olds appeared to be more aware of tempo change than other age groups, in that they found a greater dissimilarity between the standard and variations one and two than did other age groups.

Variations five and three were judged as being more similar to the standard than variations six and four, implying that an increase in the complexity of melody or rhythm had less impact, subjectively, than an objectively equivalent decrease in melodic or rhythmic complexity.

Taking 29 musically trained and 29 untrained subjects: presence or absence of musical training had no effect on similarity ratings except in the case of variations in which melodic complexity had been altered (variations five and six). In this instance, subjects with no musical training found the variations more similar to the standard than those who had received some musical training.

One of the most striking findings to emerge from this experiment concerns the discrepancy between responses requiring little verbal fluency and those which did require some degree of linguistic skill. As can be seen from Table



3.3.8, very few of the 7- to 8-year-olds actually gave a verbal response: 64% made no response to the second question on each page of the questionnaire, as contrasted with only 24% of the adults. About 43% of the 10- to 11- and 13- to 14-year-olds made no response, placing them between the two extreme age groups. This might on first appearance imply that the 7- to 8-year-olds had not been able to detect as many differences as the adults between the variations and the standard sequence, but this would be a mistaken supposition on two accounts: by consulting Figures 3.3.1 and 3.3.2 it can be seen that overall the similarity ratings given by 7- to 8-year-olds are not very different to those given by adults except possibly in the case of the variation whose rhythm was less complex than that of the standard, as shown in Figure 3.3.3. Secondly, the 7- to 8-year-olds, when required to compare the standard with itself under the guise of an additional variation, gave scores which were just as high as, if not higher than, those given by the adults and other age groups, as can be seen from Table 3.3.1. This suggests that the 7- to 8-year-olds were able to accurately recognise identity, as well as differences, between sequences.

It is also interesting to see that the number of "other" responses given by adults is greater than that given by other age groups (Table 3.3.8), suggesting that the range of original and metaphorical possibilities available in the realm of music description may increase with increasing age. This phenomenon was also observed in Experiments 1 and 2, although in a different context.

These findings provide strong confirmation of the suggestion made in Part 2 that verbal measures alone are an insufficient index of the actual perception of music by an individual. The manner in which subjects use language to communicate characteristics of music is nevertheless of interest in itself, and will be explored later in this thesis (Experiment 6).

### 3.4 : Experiment 4

#### The Effects of Variations in Objective Complexity on Perceived Complexity of, and Liking for, Quasi-Musical Sequences

##### Method

##### Subjects.

The subjects were 107 undergraduate students of psychology. Their ages ranged from 18 to 46, with a mean of 20.25 years. There were 71 females and 36 males.

##### Experimental design.

Subjects were divided into two groups. A between-subjects design was implemented: one group of subjects rated the complexity of seven musical sequences as compared with a standard sequence, and the other group rated how much they liked each of the seven sequences as compared with the standard. Within each of the stimulus sequences presented, objective physical complexity varied with respect either to rhythm, or tempo, or melody, only one of these parameters being varied for any one sequence. The standard was presented as a variation to be rated, in comparison with itself, as a control measure. The mean ratings for each of the seven stimulus sequences were examined to find out whether any relationship existed between rated liking and rated complexity. Liking and complexity ratings

were also investigated separately for each of the three parameters varied (ie melodic complexity, rhythmic complexity, and tempo). A 2x2 factorial design was used, the first factor being the (objectively measured) complexity of the variations. This had two levels because variations were either more complex than the standard or less complex than the standard by an equal amount, according to objective measures described in Experiment 3. The second factor was the amount of musical training or experience possessed by the subjects: again there were two levels because subjects were divided into a "highly trained/experienced" group and a group with little or no training, on the basis of information given on their response sheets.

#### Musical material and questionnaires.

The musical material was identical to that used in Experiment 3. The stimulus sequences were composed so that melodic complexity, rhythmic complexity and tempo varied one at a time with respect to a standard sequence. This gave rise to six variations, three of which incorporated an increase in tempo, melodic complexity, or rhythmic complexity with respect to the standard (as measured objectively), and the other three of which incorporated a decrease in tempo, or melodic or rhythmic complexity. A seventh "variation" was included which was in fact identical to the standard, although subjects were not informed at the outset that this was the case. The variations were presented in random order, and the tape used was the second experimental tape described in Experiment 3. Full



details of the sequences are given in the report of that experiment and in Appendices 3.3.1 to 3.3.6.

One half of the subjects were required to give preference ratings, and the other half rated the complexity of each sequence as compared with that of the standard. Each of those rating complexity received a sheet giving the following instructions:

You will hear seven short sequences of music and a standard sequence which will be used as a comparison. The first time you hear them they will appear one after the other, preceded by one appearance of the standard sequence. This section of the tape is played just to give you an idea of what the sequences and the standard are like, so you do not need to write or try to make any judgements about the sequences at this stage. The second time you hear them, each of the sequences will be preceded by the standard. Your task is to decide how complex you think each sequence is compared to the standard. The tape will be stopped after each pair of sequences (ie standard-sequence 1, pause; standard-sequence 2, pause) to give you time to make a decision. There are no wrong or right answers, this is just a question of personal opinion. Each time, when you have decided how complex the sequence is relative to the standard, please mark the relevant box on your response sheet.

Subjects rating preference received identical instructions except that the fifth sentence read, "Your task is to decide how much you like each sequence, compared to the standard.", and for this group the last sentence was, "Each time, when you have decided how much you like the sequence relative to the standard, please mark the relevant box on your response sheet.".

There followed a seven-point rating scale for each of the seven sequences, ranging from much less (1) to much more (7): these scales referred either to complexity or to liking in the two subject groups. Subjects were also asked to give details of their age and sex and to write down details of any musical training or experience they had received. A sample response sheet is shown in Appendix 3.4.1.

#### Procedure.

Subjects were divided into two groups, and it was explained that they were going to be asked for their opinions about seven sequences of music, as compared with a standard sequence, by marking down their responses on rating scales which would be distributed. They were asked to avoid conferring with one another or looking at other people's responses, and the response sheets and instructions were then given out. They were asked to read the instructions, and after they had done this the first part of the tape was played (i.e. standard, variation 5, variation 4, standard, variation 1, variation 3, variation 6, variation 2). This was introduced as the standard sequence, followed by the seven other sequences which they would be required to compare with the standard. After ensuring that the

understood the instructions, the remainder of the tape was played (i.e. standard, variation 5; standard, variation 4; standard, standard; etc.), allowing about 10 seconds between each pair of sequences for subjects to make their responses. It was stressed that they should listen to all of each sequence before making a decision. At no time were the natures of the ratings mentioned, as it was felt advisable to keep subjects unaware of the fact that one group was rating complexity and the other the extent to which they liked the sequences. After making all their ratings, subjects filled in details of sex, age and musical experience.

#### Analysis of Responses and Results

All ratings were converted to numerical values of one to seven from much less complex (or liked) to much more complex (or liked) respectively. Thus a sequence which was judged by one subject as being of the same complexity as the standard achieved a score of four, for that subject. A sequence which was liked much less than the standard by a subject in the liking condition achieved a score of one, for that subject. Mean ratings were calculated for each of the six variations, and for the standard when presented as a variation, over all subjects in each condition, giving rise to seven grand means for complexity ratings and seven for liking ratings. These are shown in Table 3.4.1. The relationship between perceived complexity and liking was examined by performing a Pearson's product moment correlation on the two sets of data. A positive correlation was



obtained ( $r = .60$ , 5 d.f., two-tailed), but this was not statistically significant ( $p > 0.10$ ).

Subjects were then divided into groups according to the extent of their musical training and experience. This was achieved by asking two independent judges to rate each subject on a five-point scale (0 = no training or experience, 4 = very extensive training and experience), basing their ratings on the information given by each subject on his/her response sheet. Inter-judge reliability was high ( $r = .88$  with 106 df, two-tailed,  $p < 0.001$ ), so the mean of the two judges' ratings was calculated for each subject. A five-point scale was used because it was intended initially to compare ratings given by subjects achieving scores of 0 and 4, but only three subjects scored 0, so responses given by subjects who achieved a mean "musical experience" rating of more than 2 were separated from responses given by subjects who achieved a mean rating of less than 2. Subjects with a mean rating of exactly 2 were allocated randomly to the more experienced and less experienced groups thus obtained, so that an equal number of subjects with a mean rating of 2 belonged to each group. For convenience the two groups will be referred to as the "experienced" and the "inexperienced" groups. Subjects' responses were separated further to give four groups: liking ratings given by experienced subjects ( $n = 23$ ), liking ratings given by inexperienced subjects ( $n = 31$ ), perceived complexity ratings given by experienced subjects ( $n = 23$ ), and perceived complexity ratings given by inexperienced subjects ( $n = 30$ ).

Responses to the "variation" which was in fact identical to



the standard were first examined, in order to provide base-line measures. It was assumed that a subject would hear that the "variation" in question was identical to the standard and would therefore rate its complexity as being the same as that of the standard, or would like it no more and no less than the standard; so that he/she would allocate to it a rating of four (same as standard). It was therefore predicted that the mean complexity rating and the mean liking rating for this sequence would not differ significantly from the hypothetical mean rating of four. To find out whether this was the case, four, two-tailed t-tests were performed comparing in turn (1) complexity ratings given by experienced subjects, (2) complexity ratings given by inexperienced subjects, (3) liking ratings given by experienced subjects, and (4) liking ratings given by inexperienced subjects, with the hypothetical population mean of four.

Mean complexity ratings did not differ significantly from the hypothetical value: for the experienced subjects, mean complexity rating = 4.04,  $t = 0.27$  with 22 df, and for inexperienced subjects, mean complexity rating = 4.10,  $t = 1.00$  with 29 df. Neither of these values was statistically significant. However, so far as liking ratings were concerned, there appeared to be a tendency for subjects to like the so-called "variation" more than the standard, although the two were in fact identical. For the experienced subjects, mean liking rating = 4.14,  $t = 1.82$  with 21 df,  $p < 0.1$ ; for the inexperienced subjects, mean liking rating = 4.29,  $t = 2.52$  with 30 df,  $p < 0.02$ . In the case of the inexperienced subjects, therefore, this

tendency was statistically significant.

Six, two-way analyses of variance were carried out as follows with repeated measures on one factor (complexity). They are summarised in Tables 3.4.2 to 3.4.7:

1) Analysis of complexity ratings, comparing sequences varying with respect to tempo (variations one and two): faster/slower than standard x experienced/inexperienced.

2) Analysis of complexity ratings, comparing sequences varying with respect to rhythmic complexity (variations three and four): more/less complex than standard x experienced/inexperienced.

3) Analysis of complexity ratings, comparing sequences varying with respect to melodic complexity (variations five and six): more/less complex than standard x experienced/inexperienced.

4) Analysis of liking ratings, comparing sequences varying with respect to tempo (variations one and two): faster/slower than standard x experienced/inexperienced.

5) Analysis of liking ratings, comparing sequences varying with respect to rhythmic complexity (variations three and four): more/less complex than standard x experienced/inexperienced.

6) Analysis of liking ratings, comparing sequences varying with respect to melodic complexity (variations five and six): more/less complex than standard x experienced/inexperienced.

The following statistically significant F-ratios were obtained: from Table 3.4.2,  $F(1,51) = 162.31$ ,  $p < 0.001$  for the repeated measures factor. The means indicated that subjects gave significantly higher ratings of complexity to the sequence which was twice as fast as the standard than to the sequence which was twice as slow as the standard.

From Table 3.4.3,  $\underline{F}(1,51) = 169.54$ ,  $\underline{p} < 0.001$ , for the repeated measures factor. The means indicated that subjects gave significantly higher ratings of complexity to the sequence which was objectively more complex than the standard, rhythmically, than to the sequence which was objectively less complex than the standard, rhythmically.

From Table 3.4.6,  $\underline{F}(1,52) = 57.71$ ,  $\underline{p} < 0.001$ , for the repeated measures factor. Examination of the means indicated that subjects preferred the sequence which was more complex, rhythmically, than the standard to the sequence which was less complex, rhythmically, than the standard.

For Table 3.4.7,  $\underline{F}(1,52) = 24.41$ ,  $\underline{p} < 0.001$ , for the repeated measures factor. Examination of the means indicated that subjects preferred the sequence which was melodically less complex than the standard to that which was melodically more complex than the standard.

The means for all six analyses of variance are represented graphically in Figures 3.4.1 to 3.4.6. In no cases were any interaction effects, or any main effects for musical training, statistically significant.

## Discussion

The first, and rather striking, finding to emerge from the results of this experiment is that apparently not all the inexperienced subjects in the "liking" condition realised that the third sequence they rated was in fact identical to the standard, because overall these subjects tended to give ratings



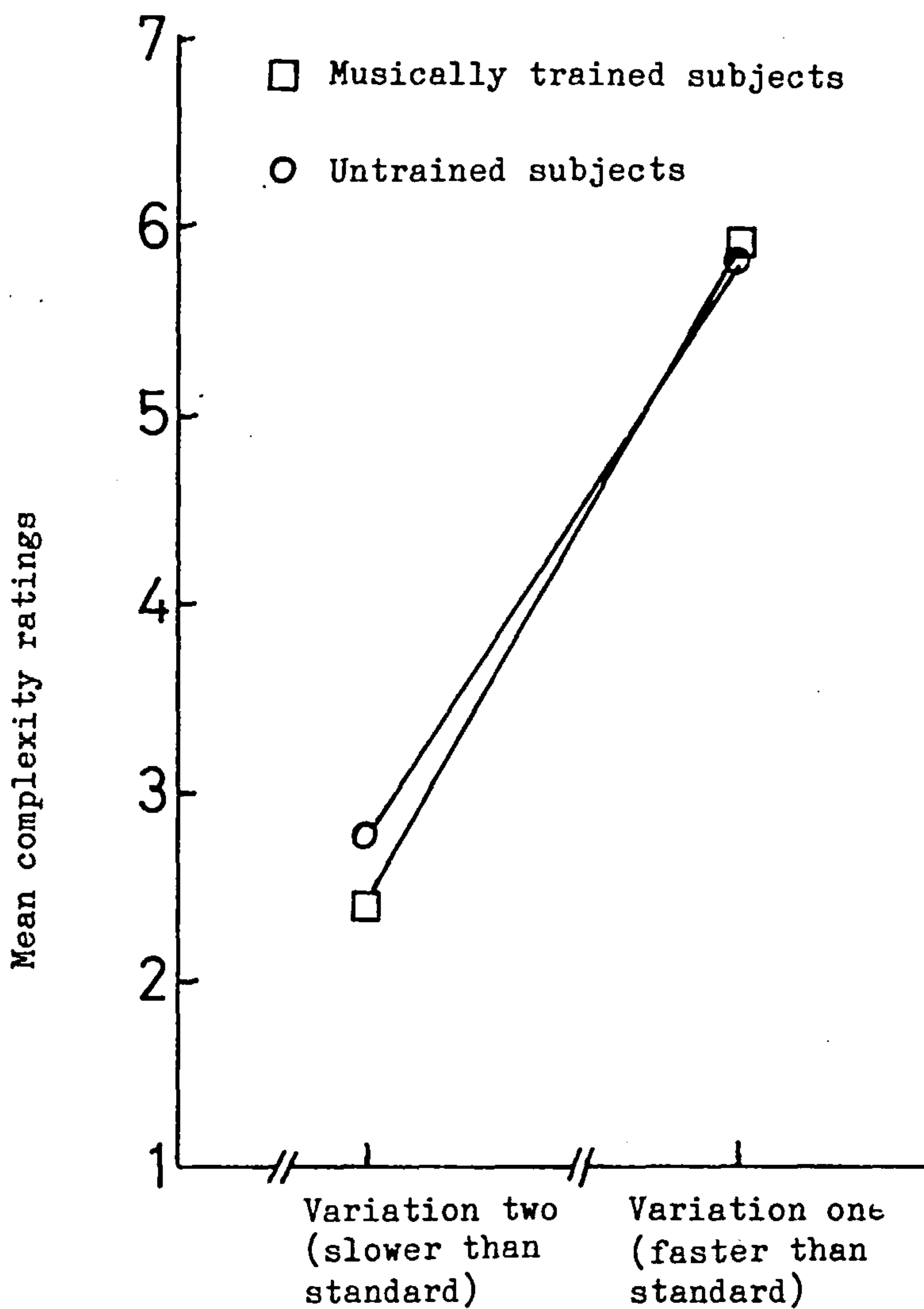


Figure 3.4.1: Mean complexity ratings for variations one and two. (On the Y-axis, 1 = much less complex than, 4 = same complexity as, 7 = much more complex than standard, respectively).



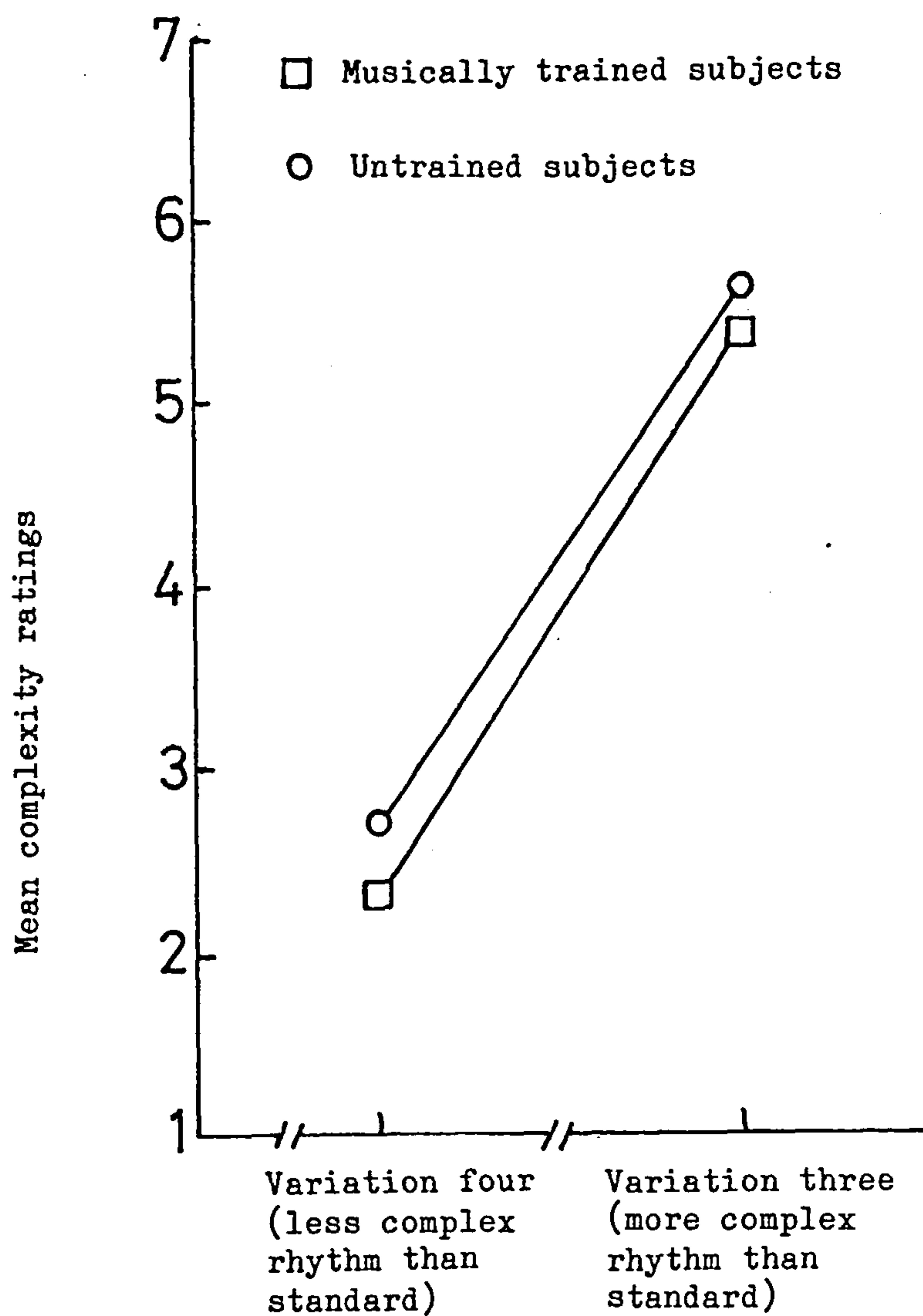


Figure 3.4.2: Mean complexity ratings for variations three and four (On the Y-axis, 1 = much less complex than, 4 = same complexity as, 7 = much more complex than standard, respectively).

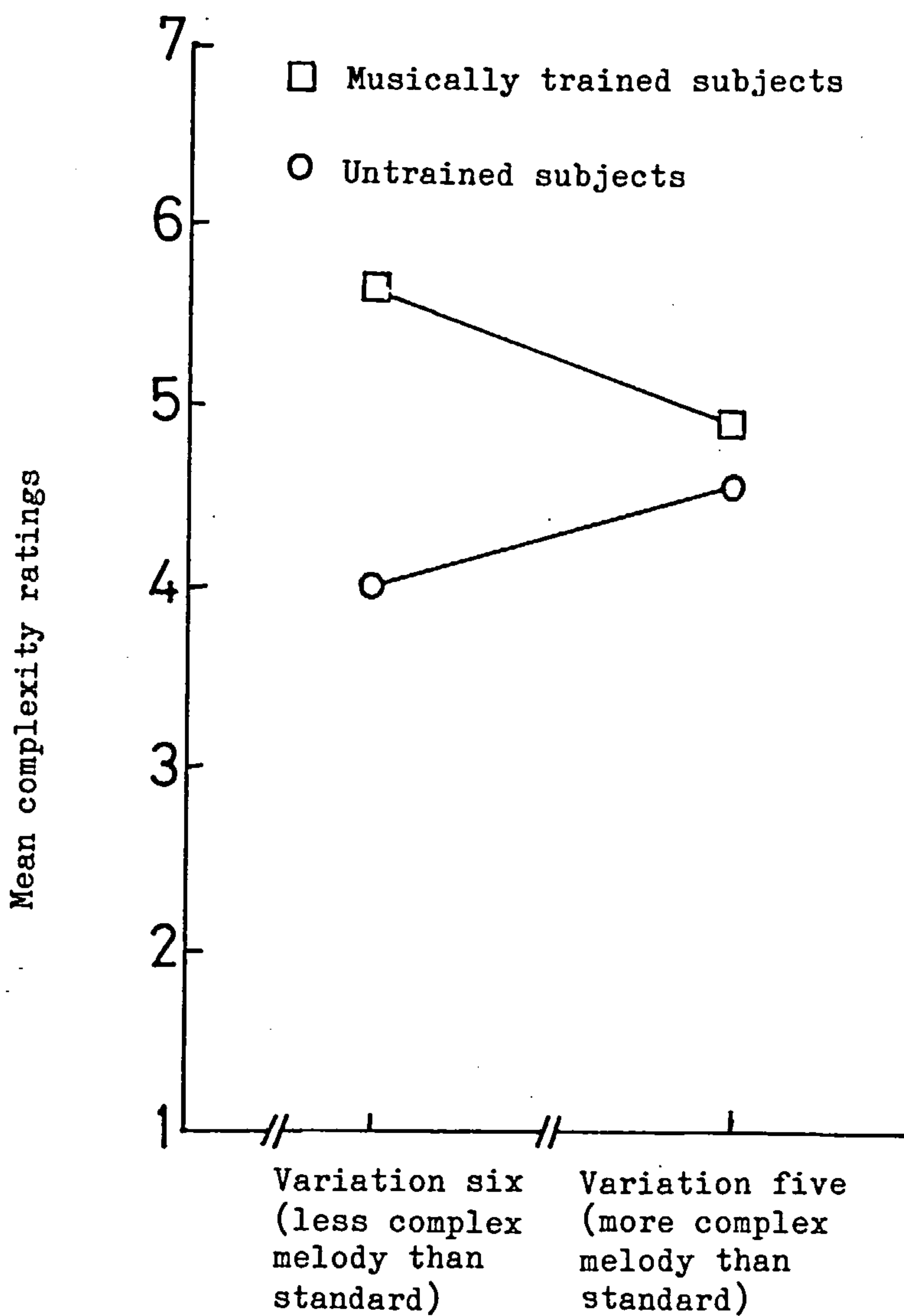


Figure 3.4.3: Mean complexity ratings for variations five and six (On the Y-axis, 1 = much less complex than, 4 = same complexity as, 7 = much more complex than standard, respectively).

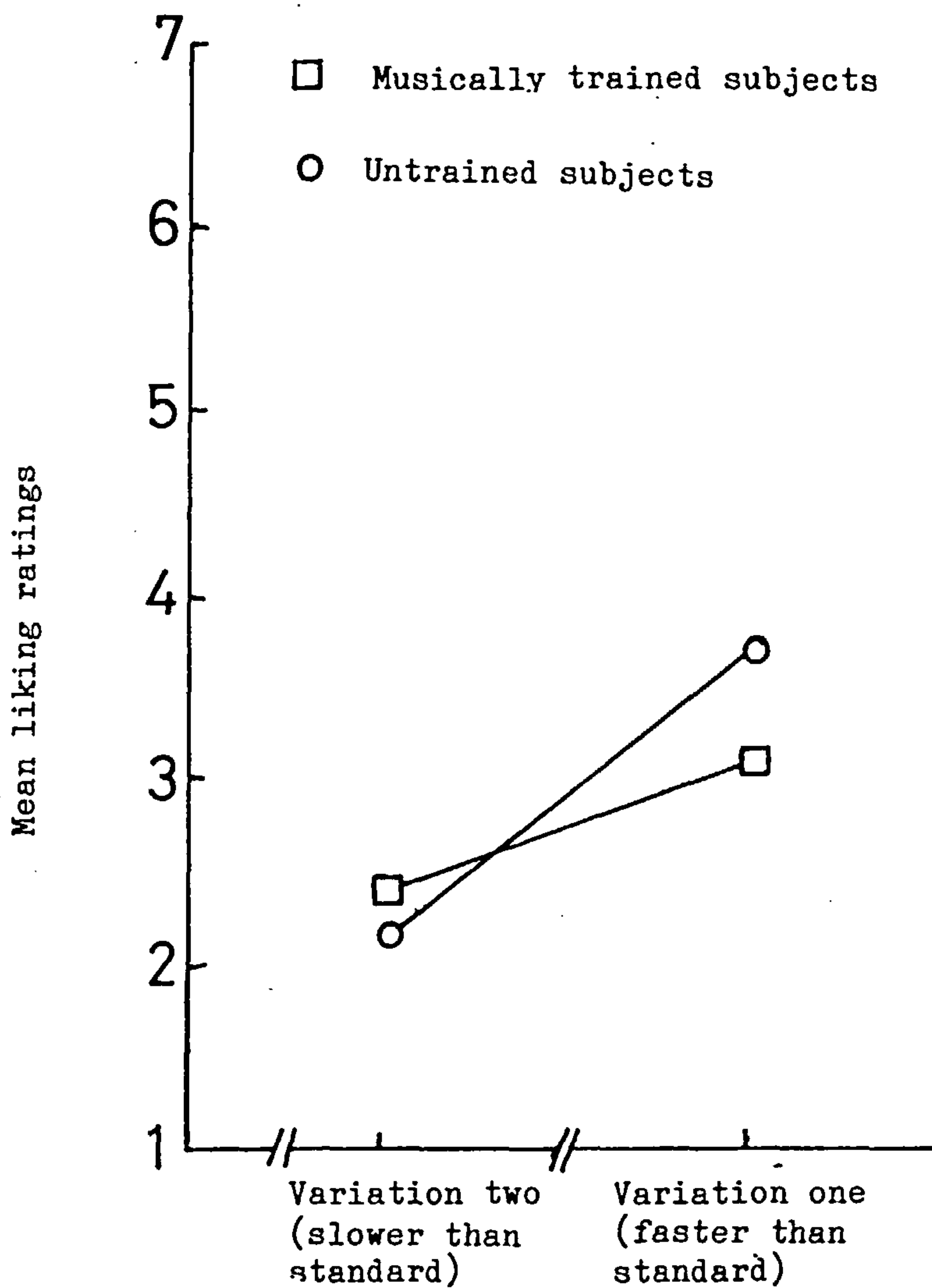


Figure 3.4.4: Mean liking ratings for variations one and two (On the Y-axis, 1 = liked much less than, 4 = liked the same as, 7 = liked much more than standard, respectively).

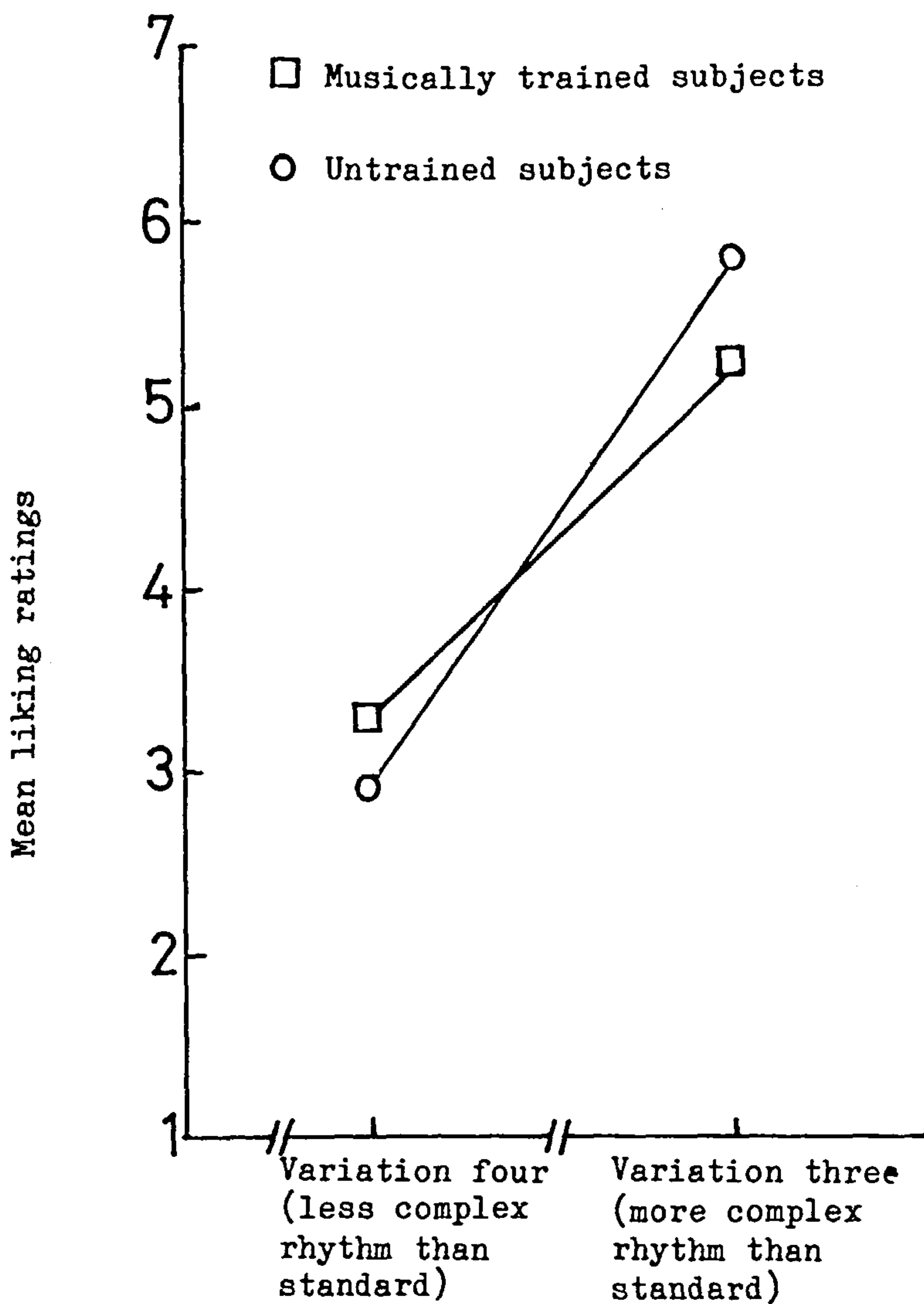


Figure 3.4.5: Mean liking ratings for variations three and four (On the Y-axis, 1 = liked much less than, 4 = liked the same as, 7 = liked much more than standard, respectively).



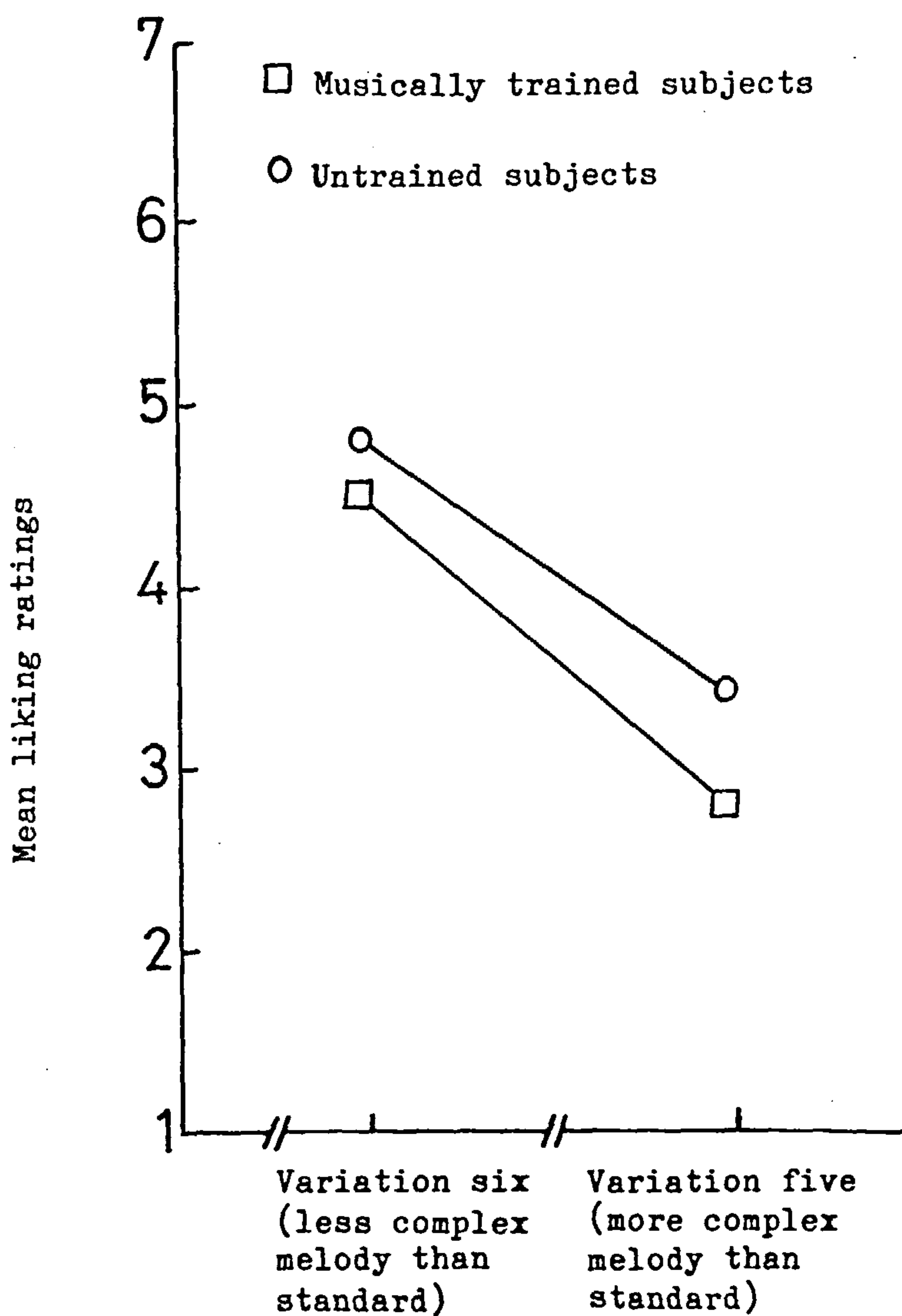


Figure 3.4.6: Mean liking ratings for variations five and six (On the Y-axis, 1 = liked much less than, 4 = liked the same as, 7 = liked much more than standard, respectively).

Table 3.4.1

Mean ratings of complexity and of liking for the seven "variations" presented as stimuli in Experiment 4.(N=107).

Type of "variation"	Complexity ratings(n=53)	Liking ratings(n=54)
s	4.07	4.22
>t	5.86	3.41
<t	2.55	2.28
>r	5.54	5.53
<r	2.51	3.10
>m	4.72	4.67
<m	4.86	3.12

Key to "variations":

- s = standard sequence (presented as a "variation")
- >t = variation one (faster than standard)
- <t = variation two (slower than standard)
- >r = variation three (more complex rhythmically than standard)
- <r = variation four (less complex rhythmically than standard)
- >m = variation five (more complex melodically than standard)
- <m = variation six (less complex melodically than standard)

Table 3.4.2

Analysis of variance summary table, n=53

The effect of musical experience on perceived complexity ratings of variations one and two relative to the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	0.580	0.83	n.s.
ERROR	51	0.701		
B	1	285.297	162.31	<0.001
AB	1	1.297	0.74	n.s.
ERROR	51	1.758		

A=musical experience, extensive or minimal (independent factor).

B=tempo, slower or faster than standard (repeated measures factor).

Table 3.4.3

Analysis of variance summary table, n=53

The effect of musical experience on perceived complexity ratings of variations three and four relative to the standard sequence.

Source of Variance	d.f.	Mean Square	F-ratio	<u>p</u>
A	1	2.199	3.70	n.s.
ERROR	51	0.594		
B	1	238.766	169.54	<0.001
AB	1	0.134	0.10	n.s.
ERROR	51	1.408		

A=musical experience, extensive or minimal (independent factor).

B=rhythmic complexity, more or less complex than standard (repeated measures factor).



Table 3.4.4

Analysis of variance summary table, n=53

The effect of musical experience on perceived complexity ratings of variations five and six relative to the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	24.112	3.32	n.s.
ERROR	51	7.264		
B	1	0.504	0.16	n.s.
AB	1	11.372	3.53	n.s.
ERROR	51	3.221		

A=musical experience, extensive or minimal (independent factor).

B=melodic complexity, more or less complex than standard (repeated measures factor).

Table 3.4.5

Analysis of variance summary table, n=54

The effect of musical experience on liking ratings of variations one and two relative to the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	0.825	0.26	n.s.
ERROR	52	3.138		
B	1	33.829	6.52	n.s.
AB	1	4.015	0.77	n.s.
ERROR	52	5.189		

A=musical experience, extensive or minimal (independent factor).

B=tempo, slower or faster than standard (repeated measures factor).

Table 3.4.6

Analysis of variance summary table, n=54

The effect of musical experience on liking ratings of variations three and four relative to the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	0.328	0.17	n.s.
ERROR	52	1.944		
B	1	155.198	57.71	<0.001
AB	1	6.309	2.35	n.s.
ERROR	51	2.689		

A=musical experience, extensive or minimal (independent factor).

B=rhythmic complexity, more or less complex than standard (repeated measures factor).

Table 3.4.7

Analysis of variance summary table, n=54

The effect of musical experience on liking ratings of variations five and six relative to the standard sequence.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	1	4.912	2.95	n.s.
ERROR	52	1.664		
B	1	63.944	24.41	<0.001
AB	1	0.389	0.15	n.s.
ERROR	52	2.619		

A=musical experience, extensive or minimal (independent factor).

B=melodic complexity, more or less complex than standard (repeated measures factor).



indicating that they liked this sequence more than they liked the standard. There was a slight tendency for this to happen in the case of the musically experienced subjects, but this did not reach statistical significance. The tendency could be a reflection of the possibility that subjects may like a piece of music more with repeated presentations. There is evidence suggesting that this may be the case under certain conditions (Rigg, 1948; Wiebe, 1940; Getz, 1966; Sluckin, Hargreaves, & Colman, 1983). The results of the present experiment might be interpreted as suggesting that the greater expertise of the experienced group enabled them to recognise identity, and to therefore resist giving any rating other than "same as standard" when asked how much they liked the third sequence, because any other response would have been a logical absurdity. It is interesting that degree of musical experience did not appear to affect complexity ratings. One explanation of this could be that the subjects rating complexity were actually not comparable with those rating liking, and that had their tasks been reversed the first group would have recognised identity even when being required to give "liking" ratings. However, care was taken to allocate subjects randomly to one group or the other, so there is no reason to suspect that the two groups differed significantly. Alternatively, an explanation could lie in the nature of the ratings themselves: although judging how complex a piece sounds and saying how much a piece is liked constitute rating procedures which are both essentially subjective, the latter type of rating intuitively seems a great

deal more subjective than the former, because in the former case it is more possible to imagine giving a wrong answer. For instance, a subject who said that a sequence sounded less complex than the standard might quite readily admit with persuasion that the sequence was in fact more complex. However, somebody who said that they liked a sequence less than the standard would not so easily be persuaded that in fact they liked it more. The suggestion is that a different approach may be employed by the listener according to whether he/she is required to rate complexity or liking. In the former case, specific stimulus attributes may be very carefully attended to in order to provide a rational basis for a decision as to how complex the music is; but in the latter case a decision may be reached on the basis of less precise characteristics such as the "feel" of the music or the associations it evokes. These results also illustrate the reason why an independent-subjects design was used. If all subjects had been rating both liking and complexity, and if it is accepted that they would be likely to recognise identity while rating complexity, then this would preclude any tendency for them to say that they liked the third sequence more than the standard (i.e. that the third sequence was not identical to the standard).

There was apparently a general trend for subjects to rate the rhythmically more complex and faster sequences as being more complex than the rhythmically less complex and slower sequences (Figures 3.4.1 and 3.4.2). Subjects rating liking for the sequences tended to prefer the former pair of sequences to the latter (Figures 3.4.4 and 3.4.5), although this trend did not

reach statistical significance when liking ratings for the faster and slower sequences were compared (Figure 3.4.4). Thus there was a tendency for the objectively more complex sequences to be rated as appearing more complex subjectively; and also for these sequences to be preferred to the less complex ones. However, the sequences in which melodic complexity had been altered constitute an exception: subjects rating complexity judged both variations to be slightly more complex than the standard, but very similar to one another with respect to perceived complexity. This is illustrated in Figure 3.4.3. Although this might suggest that subjects could not distinguish very clearly between the two sequences, examination of Figure 3.4.6 shows that this was not the case. Subjects preferred the sequence which was melodically less complex to the one which was melodically more complex than the standard, to a statistically significant extent, and this effect would be very unlikely if subjects were unable to tell the difference between the two sequences. It follows therefore that although there was a perceptible difference between them, subjects in the complexity condition did not interpret this as a difference in complexity.

In the light of the results from the liking ratings for the "variation" which was identical to the standard, and the fact that the more melodically complex sequence appeared first on the tape and the less melodically complex sequence appeared sixth, the possible existence of an order effect for liking cannot be ruled out. However, if there were a consistent order effect, with sequences appearing later on the tape being preferred to those appearing earlier, one would expect (for instance) that

variation two (appearing last) would be preferred to variation one, which appeared fourth (these were the slower and faster variations respectively). Examination of Figure 3.4.4 shows that this was not the case. It would seem more realistic to hypothesise that differences in the extent to which sequences were liked were at least partially due to differences in their objective complexity, especially since comparison pairs differed only along this dimension.

Both rhythmically and with respect to tempo the sequences sounded subjectively simple, so that any increase in the complexity of these parameters was quite likely to result in an increase rather than a decrease in liking (Heyduk, 1975). In contrast, because of the nature of the method used to generate it, the melody sounded rather unnatural even in its simplest form. The melodically more complex sequence was therefore possibly more complex than most subjects' "optimum complexity level" (Heyduk, 1975) and was therefore liked less than the standard or the less complex sequence. However, it must be remembered that all ratings were made in comparison to the standard and were therefore not absolute. In addition it would not be justifiable to assume that the standard represented the optimum complexity level with respect to melody. The point being made here is that the different pattern of results found for melodic, as compared with tempo and rhythmic variations, might be explained by suggesting that in the present context all the melodic variations were slightly more complex than the optimum level postulated by Heyduk. This would mean that an increase in complexity would lead to a decrease in liking. On



the other hand, variations of tempo and rhythm were possibly less complex than the optimum level, so that increases in complexity led to increased liking.

To return to the question of musical experience it is worth noting that for sequences where melodic complexity had been changed there was a nonsignificant trend for musically experienced subjects to allocate complexity ratings in a different way to that exhibited by inexperienced subjects. This is illustrated in Figure 3.4.3. Inexperienced subjects judged the (objectively) less complex variation to be the same as the standard with respect to complexity, and the (objectively) more complex variation to be slightly more complex than the standard. This pattern is not inconsistent with that found when tempo and rhythmic complexity were altered. However, experienced subjects rated the variations differently: they judged the (objectively) less complex variation to be more complex than the standard, and also slightly more complex than the (objectively) more complex variation. This in turn was perceived as being more complex than the standard. These trends are interesting in that they reflect to some extent the finding in the previous experiment where melodic complexity was the only parameter for which musical training influenced the magnitude of similarity judgements subjects made. In Experiment 3, subjects with musical training perceived alterations in melodic complexity as giving rise to a worse match with the standard than did subjects without training. In the present experiment, experienced subjects gave ratings suggesting that both variations differed more in complexity from the standard than was suggested by the

ratings of the inexperienced subjects (Figure 3.4.3); and the same was true, although to a lesser extent, where the liking ratings were concerned (Figure 3.4.6). Results from both experiments might therefore be taken as implying that subjects with musical training or experience are more aware of changes in melodic complexity, or perceive them as being more important as regards the general character of a composition, than are inexperienced subjects. However, neither group necessarily recognises that the parameter being varied is melodic complexity. In the present experiment these tendencies were not statistically significant and it would therefore be unwise to attach too much importance to them.

One of the primary aims of this experiment was to throw more light on the findings of Experiment 3 by discovering whether objectively measured increases or decreases in complexity were actually perceived as such. With the exception of melodic complexity, this hypothesis was borne out: variations in which rhythmic complexity or tempo had been increased (or decreased) with respect to the standard were perceived as being more (or less) complex than the standard. Changes in melodic complexity were not clearly reflected in subjects' complexity ratings of the relevant variations, but differences between liking ratings suggested that subjects had perceived some dissimilarities between them.

In Experiment 3 it was suggested that an increase in complexity (melodic or rhythmic) relative to the standard was not as apparent as the corresponding decrease in complexity. If this were reflected in the results of the present experiment, it

would mean that complexity ratings for the less complex sequences would fall further below four than the ratings for the more complex sequences rose above four. Although no statistical tests were performed specifically to ascertain whether this was the case, examination of Figures 3.4.1, 3.4.2 and 3.4.3 gives no reason to suggest that it was.

In summary, as regards the validation of the stimulus material used in Experiment 3, it would appear that an increase or decrease in complexity, measured objectively, may be perceived as such, so far as manipulations of tempo or rhythm are concerned. The case is not so clear cut as regards melodic complexity; although present evidence suggests that subjects were able to perceive a difference between a more complex and less complex melody, this was not interpreted as a difference in complexity. It would be interesting to establish exactly how it was interpreted, but the present experiment did not furnish sufficient data to allow this.

The results also suggested that a musical discrimination task may be approached quite differently depending on the type of judgement subjects are required to make. Intuitively this is not at all surprising, but further research in this area would be worthwhile, and the issue will be discussed again later in this thesis.

### 3.5 : Experiment 5

#### The Effects of Imagined Anger on Preference for Rhythms

##### Method

##### Subjects.

There were 120 subjects, most of whom were visitors to a University "Open Day" who volunteered to take part in the experiment. Age and sex were not primary criteria for their selection, but subjects were asked to identify themselves as male or female and to state to which of four age groups they belonged. This information showed that there were 56 males and 62 females, and that 20 subjects were under 15 years old, 64 were between the ages of 15 and 30, 27 were between 30 and 55 and seven were over 55 years old. Two subjects did not give any information about their age or sex.

##### Musical material and questionnaires.

All testing was carried out in a soundproof laboratory in the Psychology department at Leicester University. The musical material heard by each subject consisted of three, 12-bar phrases each made up of one of three melodic lines accompanied by one of three rhythmic backgrounds. The complexity of the rhythmic background was systematically varied according to both objective and subjective criteria: rhythmic complexity was



defined in terms of the number of events per bar and the position of these events (on or off the beat) in addition to an independent, subjective complexity judgement. Full details of both these measures appear in Appendices 3.5.1 and 3.5.2 respectively. Three rhythmic backgrounds, A, B, and C were recorded on a master tape. One of these (A) sounded very simple, a second (B) was of intermediate complexity, and the third (C) sounded very complex.

Three melodic lines, each 12 bars long, were also prepared. These served only to make the rhythmic aspect of the stimulus material sound as much like natural music as possible, so it was important that the melodies varied in very few respects, especially so far as their complexity was concerned. This was achieved by taking sequences of notes from Davies' statistically derived approximations to Western music (1969: also described in Davies, 1978, pp. 76-79), using sequences with a high degree of organisation (seven notes' restraint). The three melodies thus obtained (1, 2, and 3) are given in Appendix 3.5.3.

To control for the possibility that the listeners' perceptions of rhythmic backgrounds A, B and C might vary according to which of the three melodies was accompanying them, three experimental tapes were made up. Each tape contained the three degrees of rhythmic complexity, but in a different order and with different rhythm/melody combinations incorporated. Subjects hearing the first tape listened first to melody 1 accompanied by rhythm A (simple); then melody 2 accompanied by rhythm B (intermediate complexity); then melody 3 accompanied by rhythm C (complex). The composition of the second and third

tapes was: melody 1, rhythm B; melody 2, rhythm C; melody 3, rhythm A and: melody 1, rhythm C; melody 2, rhythm A and melody 3, rhythm B respectively.

Two types of questionnaire were produced; on the first subjects were asked to listen to the three sequences and say which, if any, they liked best; the second required subjects to imagine that they were very angry or highly annoyed and to say which sequence, if any, they would prefer to listen to in this mood. On both questionnaires subjects were given the option of saying that they would prefer silence to any of the three sequences. Subjects were also required to give reasons for their preference (if possible), to give some information about the type of music to which they liked to listen at home, to say whether they were feeling "very calm" "slightly irritable" or "very irritable" (regardless of whether or not they were imagining annoyance) and to indicate their sex and age group. The two questionnaires appear in full in Appendices 3.5.4 and 3.5.5.

#### Procedure.

Subjects were tested in groups of two, three or four in a soundproof booth in the Psychology department at Leicester University. It was explained to all subjects that they would hear three short sequences of "artificial" music played on synthesizers and then tape-recorded. Half the groups of subjects were given Questionnaire 1 (Appendix 3.5.4), the other half Questionnaire 2 (Appendix 3.5.5). Those receiving Questionnaire 1 constituted the control group, subjects

completing Questionnaire 2 were the experimental group. Members of a group in any one testing session always completed the same type of questionnaire and were not aware, throughout the duration of the experiment, that there was an alternative condition. Those receiving the first questionnaire were asked to listen carefully to the sequences and indicate which of the three they liked best, or whether they would have preferred silence, by placing a cross in the relevant box on the questionnaire. Those receiving Questionnaire 2 were asked to imagine that they were feeling very angry or highly annoyed, perhaps using the memory of a previous occasion when they had felt like this to make the image more vivid. They then listened to the sequences and were asked to say which they thought they would prefer when in this mood, or whether they would have preferred silence; again responses were given by marking the relevant box on the questionnaire. All subjects in both conditions were asked to give reasons for their preference, if they could, and to complete the remainder of the questionnaire. No time limit was set, but most testing sessions lasted for between 5 and 10 minutes. After the response sheets had been collected, the aim of the experiment was discussed with any subjects who expressed an interest.

Testing not only involved the alternation of Questionnaires 1 and 2 between sessions, but also, for any one questionnaire, rotation of the three stimulus tapes (again, between not within sessions). There were therefore six possible different combinations of stimulus tapes and questionnaires, and care was

taken to ensure that equal numbers of subjects took part under each combination. These six combinations were: (a) Questionnaire 1, tape 1; (b) Questionnaire 2, tape 1; (c) Questionnaire 1, tape 2; (d) Questionnaire 2, tape 2; (e) Questionnaire 1, tape 3; (f) Questionnaire 2, tape 3. Appendix 3.5.6 summarises the presentation of questionnaires and stimulus material.

### Analysis of Responses and Results

As regards Hypotheses 1 to 6 which were stated in Part 3.2, the frequencies with which each type of preference (i.e. for rhythmic variations or for silence) occurred were apparent upon inspection of the response sheets. The frequencies with which preferences occurred, for control and for experimental subjects, are shown in Figures 3.5.1 to 3.5.6: these refer to Hypotheses 1 to 6 respectively. As regards Hypothesis 7, taking both control and experimental groups, the reason each subject gave for his or her preference was placed in one of four categories depending on whether it involved: (a) rhythm (e.g. "There's a lot going on especially the drums" or "It's jerky and has uneven rhythm"), (b) melody (e.g. "Unusual intervals" or "Good melody"), (c) affective responses (e.g. "The third piece was very disturbing" or "More peaceful"), or (d) other responses (e.g. "Silence is a rare commodity in my surroundings"). Figure 3.5.7 shows the frequencies with which responses fell into each of the cells.

Each of the seven hypotheses was tested by means of Chi-squared tests and analytical comparisons as described in Keppel and Saufley (1980). Because a large number of tests was carried



out on different configurations of the same data, it was anticipated that some comparisons would yield statistically significant results by chance alone. Findings should therefore be treated as suggestions rather than as firm conclusions. Table 3.5.3 gives the values of Chi-squared and analytical comparisons (two-tailed tests) associated with each Figure.

### Discussion

The purpose of this experiment was to ascertain the extent to which rhythmic complexity in music affects preference decisions, and also to establish whether asking subjects to imagine they were angry or annoyed would bring about a change in the pattern of preferences. It was therefore important that the melodic components of the stimuli should be equally well liked regardless of which rhythmic accompaniment was heard with them; if one melody was consistently preferred any effect of rhythmic complexity might be masked.

In order to check this, response sheets were examined to find out the frequencies with which each melody had been selected (each subject had heard each melody once). A Chi-squared test was then performed on the frequency data (Figure 3.5.6) for preferences for each melody (regardless of rhythmic accompaniment). The value of Chi-squared did not reach statistical significance (Table 3.5.1). It was therefore assumed that any subsequent notable preference patterns emerging among the stimuli were more likely to be a reflection of variations in the complexity of the rhythms than of variations among the three

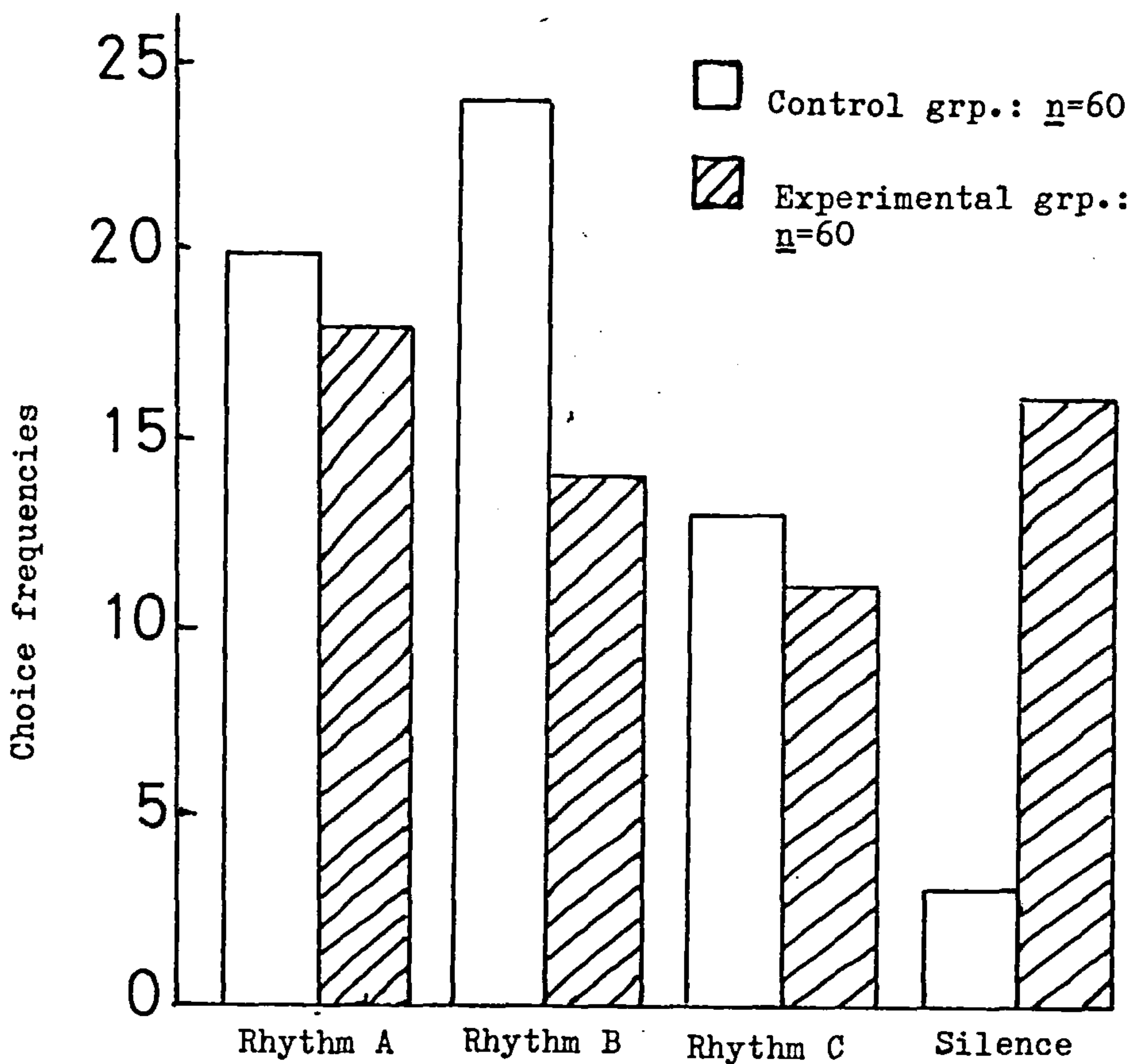


Figure 3.5.1: Frequencies with which rhythms A, B, C, or silence were preferred among the control and experimental groups. (Rhythms A, B, and C were of low, intermediate, and high complexity respectively.)

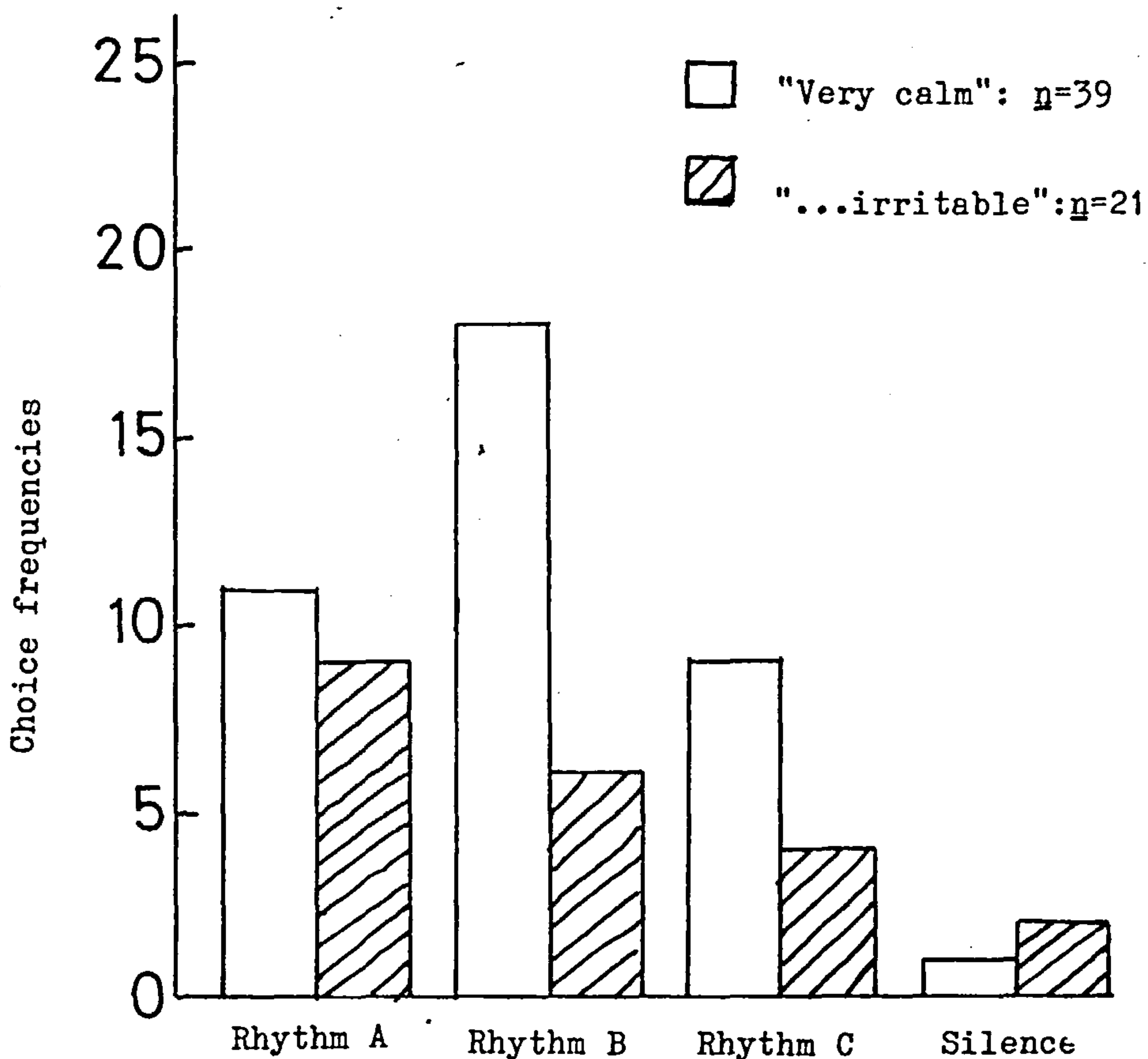


Figure 3.5.2: Frequencies with which rhythms A, B, C, or silence were preferred by subjects in the control group, for those who were "very calm" and those who were "slightly irritable"/"very irritable" by self-report. (Rhythms A, B, and C were of low, intermediate, and high complexity respectively).

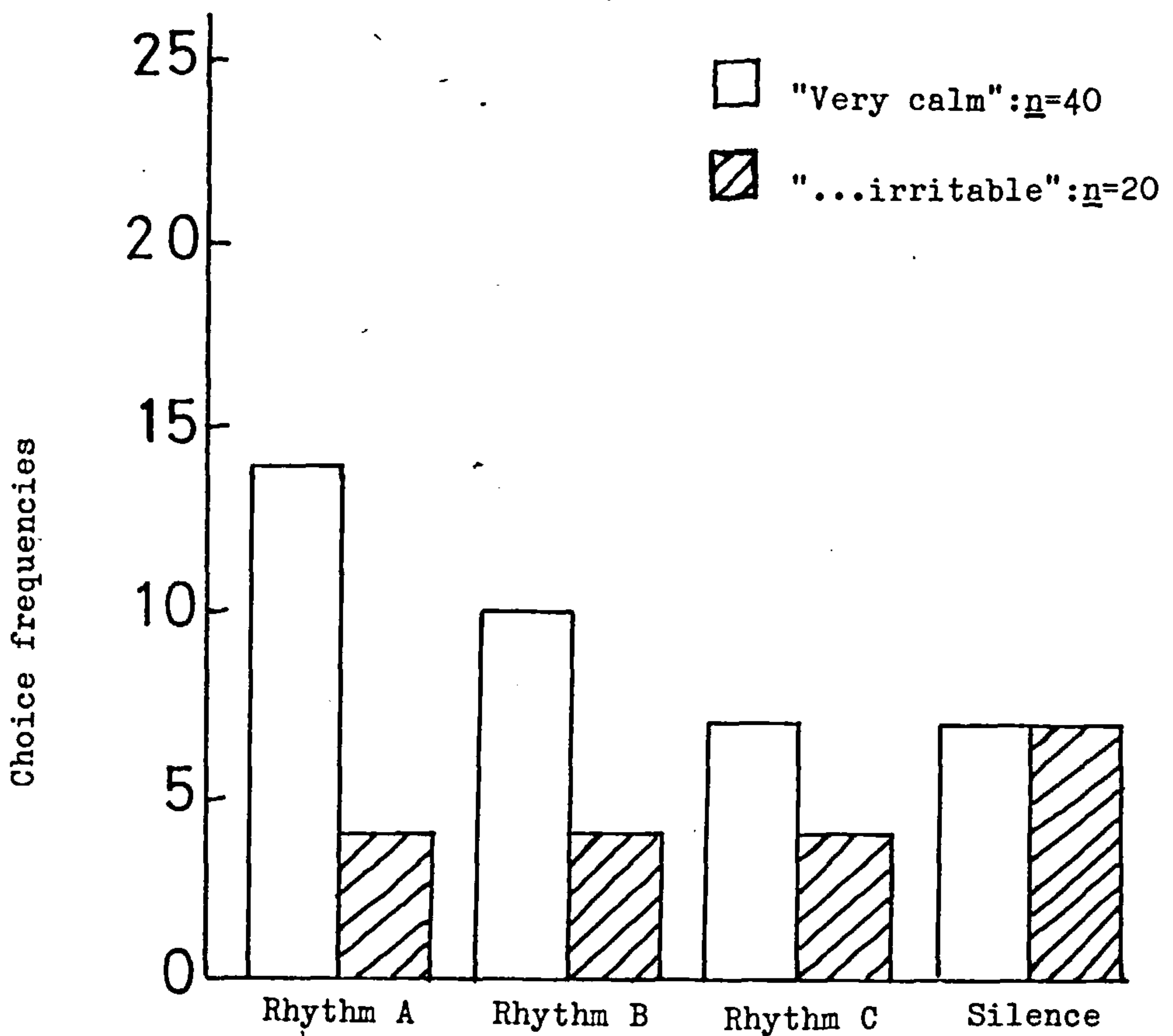


Figure 3.5.3: Frequencies with which rhythms A, B, C, or silence were preferred by subjects in the experimental group, for those who were "very calm" and those who were "slightly irritable"/"very irritable" by self-report. (Rhythms A, B, and C were of low, intermediate, and high complexity respectively).



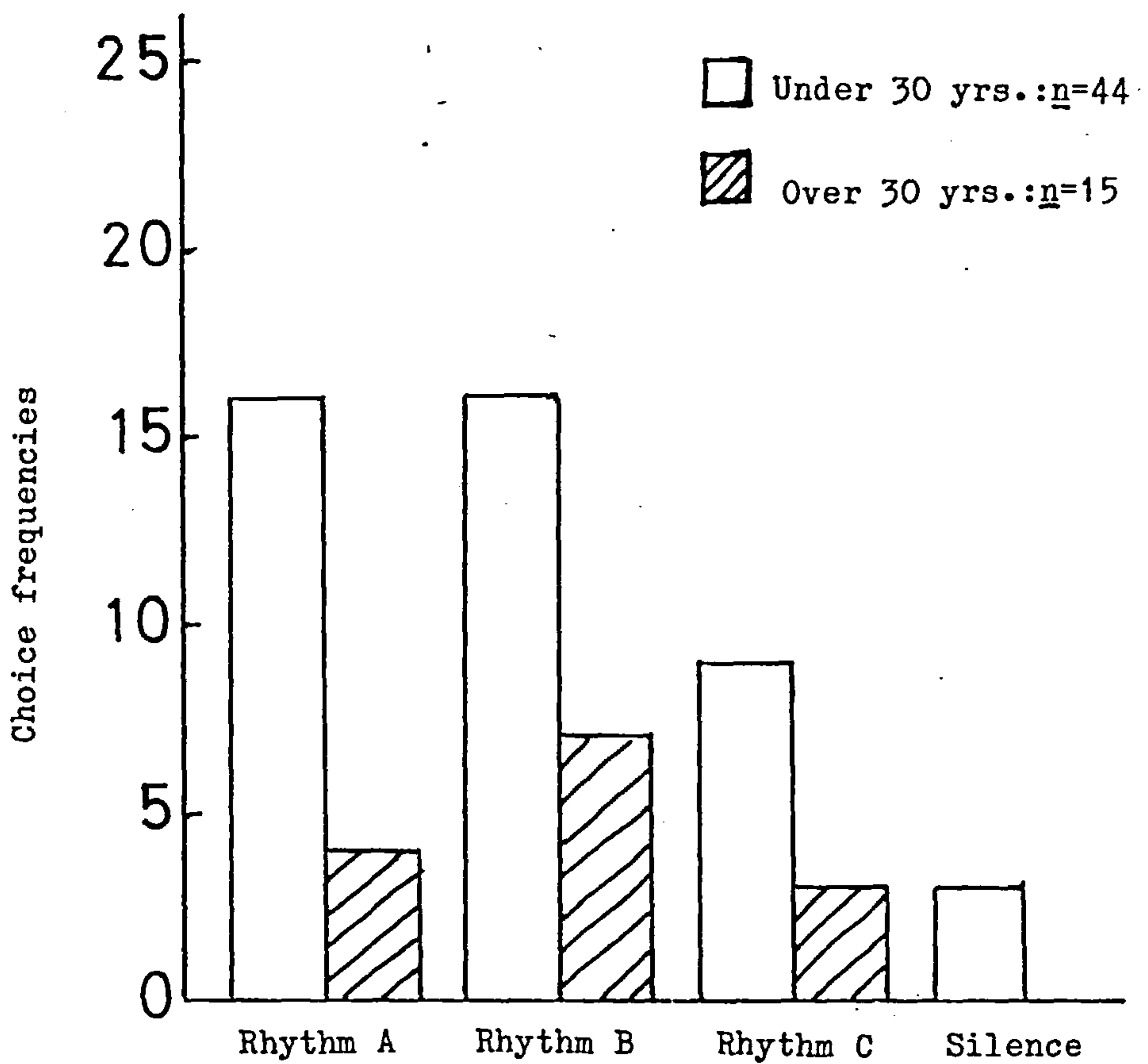


Figure 3.5.4: Frequencies with which rhythms A, B, C, or silence were preferred by subjects under and over 30 years old in the control group. (Rhythms A, B, and C were of low, intermediate, and high complexity respectively).

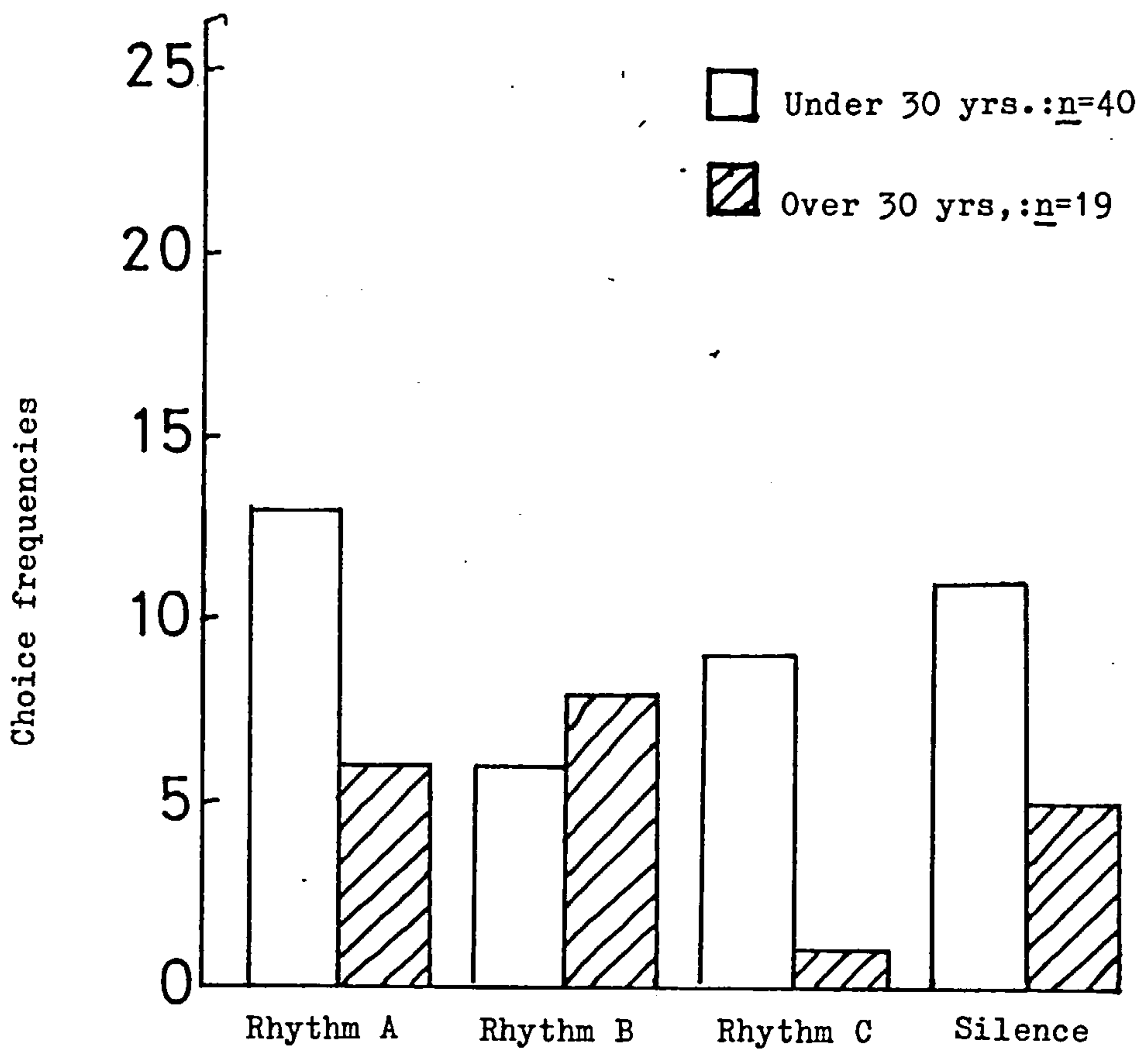


Figure 3.5.5: Frequencies with which rhythms A, B, C, or silence were preferred by subjects under and over 30 years old in the experimental group. (Rhythms A, B, and C were of low, intermediate, and high complexity respectively).

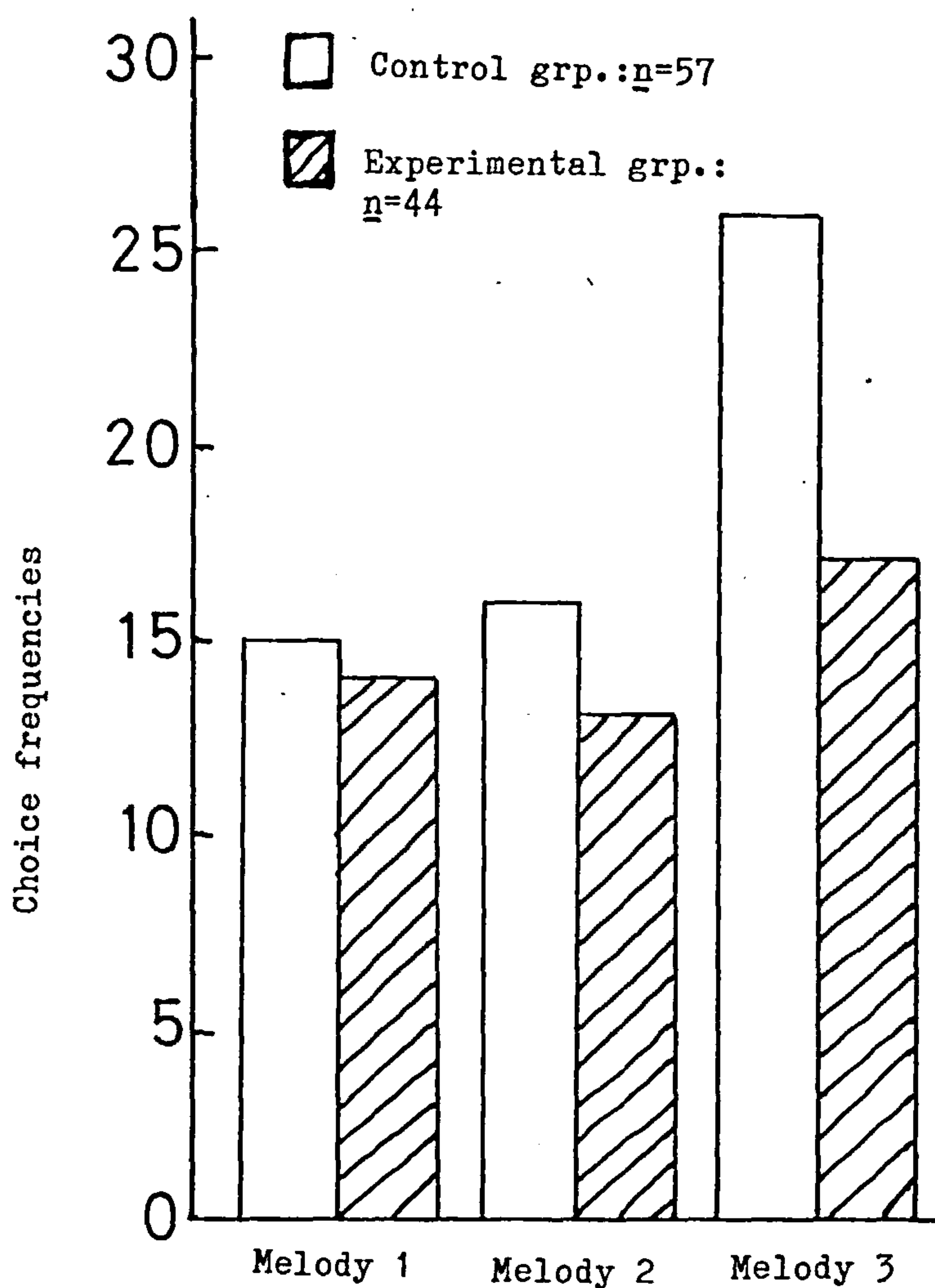


Figure 3.5.6: Frequencies with which melodies 1, 2, and 3 were preferred by subjects in the control and experimental groups. (Data from subjects preferring silence are excluded).

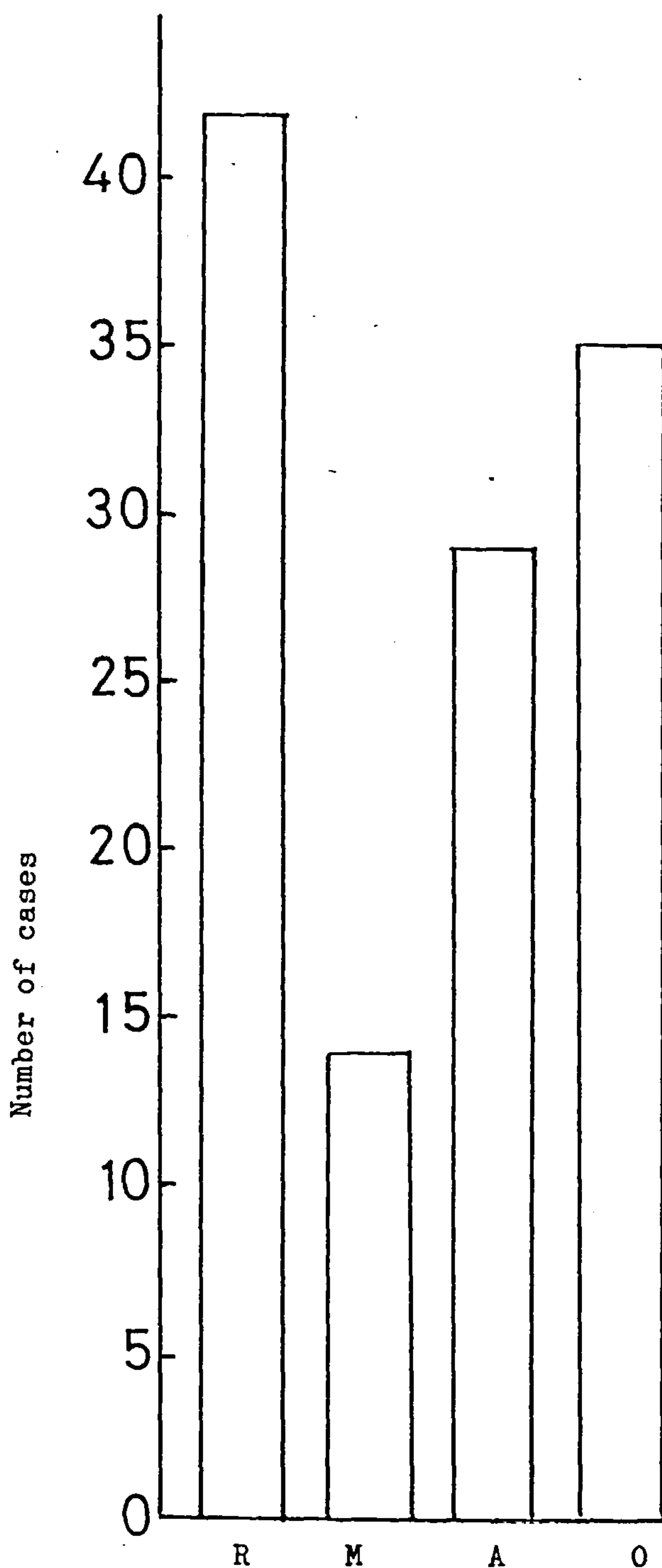


Figure 3.5.7: Frequencies with which subjects' reasons for their preferences involved rhythm (R), melody (M), an associative or affective component (A), or none of these (O).



Table 3.5.1

Results of Chi-squared tests and analytical comparisons. The numbers 1 to 7 in the left-hand column refer to Hypotheses 1 to 7 which were outlined in Part 3.2.

Details of samples/dimensions involved in test		$\chi^2$	$\chi^2$ comp.	d.f.	p	<u>n</u>
1.	Control/experimental groups x A/B/C/silence( <u>n</u> =120)	11.72		3	<0.01	
	(A+B+C)/silence		10.57	2	<0.01	
	CONTROL GROUP: (A/B/C)		4.37	2	n.s.	60
	EXPERIMENTAL GROUP:(A/B/C)		2.10	2	n.s.	60
2	CONTROL GROUP (a) "very calm" subjects: A/B/C/silence	15.05		3	<0.01	39
	"very calm" subjects: (A+B+C)/silence		10.47	1	<0.01	
	(b) "slightly" or "very irritated" subjects: A/B/C/silence	5.30		3	n.s.	21
	"slightly" or "very irritated" subjects: (A+B+C)/silence		2.68	1	n.s.	
3.	EXPERIMENTAL GROUP (a) "very calm" subjects: A/B/C/silence	2.60		3	n.s.	40
	"very calm" subjects: (A+B+C)/silence		0.53	1	n.s.	
	(b) "slightly" or "very irritated" subjects: A/B/C/silence	2.40		3	n.s.	20
	"slightly" or "very irritated" subjects: (A+B+C)/silence		2.40	1	n.s.	

Table 3.5.1(continued)

Results of Chi-squared tests and analytical comparisons. The numbers 1 to 7 in the left-hand column refer to Hypotheses 1 to 7 which were outlined in Part 3.2.

Details of samples/dimensions involved in test	$\chi^2$	$\chi^2$ comp. d.f.	p	<u>n</u>
4. CONTROL GROUP				
Subjects under 30 years: A/B/C/silence	10.72	3	<0.05	44
Subjects over 30 years: A/B/C/silence	8.74	3	<0.05	15
5. EXPERIMENTAL GROUP				
Subjects under 30 years: A/B/C/silence	2.00	3	n.s.	40
Subjects over 30 years: A/B/C/silence	4.37	3	n.s.	19
6. CONTROL GROUP				
Melody1/Melody2/Melody3	3.89	2	n.s.	57
EXPERIMENTAL GROUP				
Melody1/Melody2/Melody3	0.59	2	n.s.	44
7. CONTROL & EXPERIMENTAL GPS				
Frequencies with which reasons for responses involved rhythm/melody/ affective & associative connotations/other	14.19	3	<0.01	120
Frequencies with which reasons for responses involved (rhythm+affective & associative connotations +other)/melody	11.37	1	<0.001	

Slashes are used to separate the cells whose choice frequencies are being examined by the test in question.

melodies. This hypothesis is further supported by examination of subjects' comments about the three musical stimuli (Figure 3.5.7): when asked what determined their preferences, only 14 out of 120 (11.7%) mentioned melodic aspects of the music, as contrasted with 35% whose responses involved rhythm. 24.2% reported that they had made their judgements on the basis of affective/associative connotations, and the remaining 29.1% gave no reason or reasons which fell into no other categories.

From Figure 3.5.1 it can be seen that the patterns of preference for the control and experimental groups differed. Out of the control group, 40% preferred pieces having rhythm B (intermediate complexity), 33.3% preferred pieces with rhythm A (simple), and rhythm C (complex) was least frequently preferred: only 21.7% made this choice. Out of the control group, 5% would have preferred silence. By contrast, among the subjects imagining themselves to be angry, 31.7% liked the simple rhythm best and 23.3% preferred the rhythm of intermediate complexity. Rhythm C (complex) was again least often chosen (by 18.3% of subjects); and in this case, 26.7% of the subjects said that they would have preferred silence.

This pattern of results is consistent with findings such as those by Berlyne (1971, 1974) in which the relationship between musical complexity and liking was found to conform to an inverted-U curve with liking first increasing and then decreasing as complexity increased. The control group in the present experiment show a preference pattern which could be said to demonstrate this phenomenon with respect to rhythmic complexity (Figure 3.5.1). Konecni and Sargent-Pollock (1976)

pointed out, "An important derivative of Berlyne's model (e.g. 1967) is that the choice of stimuli in the higher ranges of collative dimensions (eg relatively complex patterns) should decrease when the level of arousal is high." (p. 347). This view of the situation hinges around the assumption that a high level of arousal will cause subjects to prefer lower levels of complexity in any type of stimulus. Konecni (1979) tested this hypothesis with respect to melodic complexity and found support for it in that subjects who had been made angry (and hence had a higher level of arousal than a control group) preferred simple to complex melodies. In the present experiment the experimental group showed a preference for the simple rhythm, with liking for remaining pieces decreasing with increasing rhythmic complexity (Figure 3.5.1). It is interesting that this trend emerged even though the experimental group in this case was not actually angry but merely imagining the mood. For both control and experimental groups, trends among rhythms were not statistically significant. However, results of the Chi-squared tests showed that although the control group was extremely unlikely to choose silence, the experimental group was as likely to prefer silence as any of the three types of rhythm. This again is not inconsistent with the notion that subjects with a high level of arousal, and hence, according to Konecni and Sargent-Pollock (1976) and Konecni (1979) a reduced processing capacity, are more likely than a control group to choose stimuli which demand



little or no processing effort.

Considering the control group alone (Figure 3.5.2) it was found that subjects who described themselves as "very calm" were more likely to choose one of the three sequences than silence, whereas subjects who were by self-report "slightly irritated" or "very irritated" were equally likely to choose silence or a musical stimulus. This effect was statistically significant. Examination of the histograms in Figure 3.5.2 shows that subjects describing themselves as "very calm" had a similar preference pattern to that of the control group as a whole (Figure 3.5.1), while the preferences of those describing themselves as "slightly irritated" or "very irritated" followed a similar pattern to those of subjects who were imagining anger (Figure 3.5.1, experimental group). This suggests that if subjects were either imagining anger or feeling "slightly irritated" (the latter being a previously existing mood not intentionally induced by experimental procedure) the effect was the same: they were as likely to prefer silence as they were to prefer any of the stimulus sequences.

From Figure 3.5.3 and the associated values of Chi-squared, it can be seen that all subjects in the experimental group (i.e. all those imagining that they were angry) showed the same pattern of preference whether their actual mood at the time was "very calm", "slightly irritated", or "very irritated": all were as likely to choose silence as music. There appeared to be a tendency for subjects who were imagining anger and were in addition actually irritated to choose silence more frequently than music, but this was not statistically significant.

Overall, Figure 3.5.3 suggests that any difference in preference patterns that might have occurred due to subjects' own moods, as seen in the control group, was masked by the task requirements of the experimental group.

Figure 3.5.4 shows the distributions of preferences for control group subjects under and over 30 years old. The older subjects tended to favour rhythm B which was interesting because this rhythm closely resembled the "swing" music fashionable when many of these subjects had been in their teens. However, there were only 15 subjects in the control group who were over 30 and this effect was not statistically significant. Distributions of preference for the two age groups did not differ significantly for the experimental subjects (Figure 3.5.5).

In summary, it emerged that subjects pretending to be angry, and subjects who actually were irritated, were as likely to prefer silence as they were to prefer any of the three rhythmic sequences; whereas those in the control group who were "very calm" by self-report preferred one of the three sequences to silence. No statistically significant findings emerged with respect to differences in preference among the three rhythmic sequences A, B, and C, but any trends that were observed were consistent with previous research (e.g. Konecni 1979) in that subjects with a higher arousal level (albeit imaginary) and hence, according to Konecni, less processing capacity tended to choose the simplest rhythm or silence most often (Figures 3.5.1 and 3.5.3). Subjects with low levels of arousal (control group, "very calm" by self-report) tended to prefer the rhythm of intermediate complexity (Figure 3.5.2). These findings may be

brought together under Walker's (1973) notion of "psychological complexity" mentioned in Part 3.1: although the control and experimental groups both heard the same three degrees of (objective) rhythmic complexity, A, B, and C, it is possible through Walker's approach to conclude that the psychological complexity of the sequences was greater for the experimental group than the control group, and this would account for the different patterns of preferences observed in the two groups.

One further point which should be made as regards the generality of findings from any research in which a rhythm synthesizer is used is that, as Gabrielsson (1982) remarked, "...musicians do not play with metronomic regularity or like machines." (p. 42). Thus, results from any research in which a rhythm of "metronomic regularity" is involved may not be relevant to "real-life" reactions to music. Indeed, many subjects remarked on the "clockwork" nature of the sequences. However, even if the rhythm in the present experiment deviated somewhat from that characteristic of naturalistic music in its inflexibility, it could be argued that this deviation is no greater than the variations found among actual performances of natural music: Gabrielsson (1982) found that similarly notated rhythmic patterns were actually played quite differently in different musical contexts.

### 3.6 : Conclusions (Part 3)

The aims of Experiment 3 were to assess the extent to which similarity ratings for sequences of experimental music were affected by age and by the amount of musical experience subjects possessed. Experimental musical stimuli were used so that any alterations in subjects' responses could be traced to alterations in specific physical stimulus characteristics; and a subsidiary aim of the experiment was to determine whether subjects' perceptions of the magnitude of alterations in complexity concurred with objective measures of the same.

The results of this experiment were interesting in that there appeared to be no systematic developmental trends in subjects' ability to perceive differences in melodic complexity, rhythmic complexity, or tempo. Although some age-related effects were found, none of them were such that it could be argued that the younger subjects were less sensitive to changes within the stimulus material than were the older subjects. These findings contrast with those of Experiments 1 and 2 where older subjects, particularly adults, were more sensitive to stylistic differences than were younger subjects. The explanation for this apparent contradiction probably lies in the nature of the tasks set during the experiments. In Experiments 1 and 2, subjects had only two response options and their responses had to be expressed verbally: they had to say whether two excerpts came from the same or from different sources. In Experiment 3 the range of responses was less limited, and



responses were expressed nonverbally: subjects indicated degree of similarity by circling one out of ten possible options. Both the visual arrangement of these options (a low degree of similarity on the left, increasing to a perfect match on the right), and the familiarity of the response method used (giving marks out of 10) were probably instrumental in helping the younger subjects to understand what was required of them. The important implication of these findings is that children even as young as 7 or 8 years old do not necessarily perform less well than adults on tasks of the kind described, although it may be difficult to find appropriate methods of measuring their ability.

Musical experience did not appear to affect subjects' ability to detect alterations in the tempo or rhythmic complexity of the stimuli presented. However, when melodic complexity was manipulated, musically experienced subjects gave lower scores in terms of degree of match to a standard than did inexperienced subjects. This indicates that the former group may have been more aware of changes in melodic complexity than were the latter group, which was what might be predicted on the basis of Conley's (1981) findings.

Subjects in Experiment 3 did not, on the whole, recognise that the magnitudes of the two manipulations of rhythmic complexity, or those of the two manipulations of melodic complexity, were equal. In both cases, they rated the "less complex" variation as being less similar to the standard than the "more complex" variation. On the other hand, they gave ratings to the variations in which tempo had been altered which

indicated that they perceived both variations as being equally dissimilar to the standard. These findings make it clear that objective and subjective measures of changes in musical stimulus characteristics do not necessarily correspond, so it might be advantageous in future research to take both types of measure into account.

The relationship between objective and subjective measures of complexity was examined in Experiment 4, in which one group of subjects was asked to rate the perceived complexity of each sequence used in Experiment 3, relative to the standard. It was anticipated that the variations which were faster, or objectively more complex, rhythmically and melodically, than the standard would be perceived as more complex than the standard; and that the "variation" which was in fact identical to the standard would be perceived as possessing a degree of complexity equal to that of the standard. It was predicted that the remaining three variations would be perceived as being less complex than the standard. All these predictions were borne out, with the exception of those regarding the variations in which melodic complexity was manipulated. In this case, both variations (ie those objectively less complex and objectively more complex than the standard) were perceived as being slightly more complex than the standard, and there was no significant difference between the complexity ratings received by each. This finding adds further support to the suggestion that objective and subjective measures of musical complexity may not necessarily correspond, at least where melodic complexity is concerned.

Musical experience did not seem to play an important part in determining the extent to which subjects perceived the variations as being more, or less, complex than the standard, because none of the effects produced by comparing musically experienced with inexperienced subjects reached statistical significance. On examination of affective ratings given by a second group of subjects, the same phenomenon was observed with respect to the amount the variations were liked relative to the standard: amount of musical experience was apparently not important in determining degree of preference.

However, the extent to which subjects liked the variations relative to the standard was influenced by the objective complexity of the variations, relative to that of the standard. Where melodic complexity was the parameter under consideration, the less complex variation was preferred to the more complex variation: but where rhythmic complexity and tempo were concerned, the more complex variation was preferred. (In the case of tempo, this trend was not statistically significant.) These findings suggested that subjects were in fact able to detect some difference between the two variations in which melodic complexity had been altered (if they had noticed no difference, then one variation would presumably not have been preferred to the other). However, they did not consider that this difference was a difference in complexity, as revealed by their complexity ratings; so it would appear that musical preference may be influenced by melodic complexity without subjects being aware of exactly what factor is influencing their choice.

The influence of complexity upon preference was examined further in Experiment 5, where it was found that variations in rhythmic complexity did not have a marked effect on preference. However, subjects who were imagining that they were angry exhibited a slightly different pattern of preference from that shown by subjects in the control condition. The former group were as likely to prefer silence as they were to prefer any of the musical stimuli, whereas the control group preferred music to silence. This finding was discussed in the light of Konecni and Sargent-Pollock's (1976) and Konecni's (1979) research, for which it offered partial confirmation.

The common feature of Experiments 3, 4, and 5 was that all of them used experimental, as opposed to naturalistic, musical stimuli. Using these stimuli, and a largely nonverbal response method, the most prominent finding was that similarity judgements did not appear to be systematically affected by age of subjects nor, with the exception of one type of stimulus manipulation, by the extent of subjects' musical training. Judgements of complexity were not affected by age, but they were influenced by changes in the (objective) complexity of the stimulus sequences heard. Preferences for the stimulus sequences did not appear to vary with amount of musical experience, but did vary with the objective complexity of the sequences; and liking was also somewhat affected by simulated anger.

Although a developmental trend was observed in Experiments 1 and 2, such that sensitivity to musical style appeared to increase, overall, between the age of 7 years and adulthood,



this finding was not supported by an examination of the developmental trends in "degree of match" judgements (Experiment 3). It has been suggested that this was because the younger subjects in the latter experiment were not put at a disadvantage by the need to give verbal responses. It would also seem likely that sensitivity to musical style (in the normally understood use of the word "style"), amounts to rather more than can be quantified by measuring sensitivity to individual and separate stimulus characteristics within music. The findings of the experiments reported so far may therefore be explained by suggesting that although there may not be clear-cut developmental trends in sensitivity to any one musical parameter individually, a subjects' ability to process a complex combination of parameters (such as is found in naturalistic music) does alter considerably with age. The remaining experiments in this thesis contain some investigation of the development of the perception of style, and the role of stylistic labelling, per se; but without any attempt to determine which stimulus features give rise to specific stylistic labels.

## PART 4 : LABELLING OF MUSICAL STYLES:

### A COGNITIVE/NATURALISTIC APPROACH

#### 4.1 : The Role of Categorisation in Music Perception

It has been argued by many authors (Quine, 1969; Rosch, 1978; Tversky & Gati, 1978) that the ability to categorise stimuli and to classify situations according to their similarity is fundamental to survival. Without this ability, to take some extreme examples, prey would be indistinguishable from predator, friend from foe, and poisonous plants would be eaten as readily as edible ones. The importance of categorisation is reflected in the structure of language itself: the fact that one word (e.g. "vehicle") may stand for a wide range of objects despite the fact that all these objects also have individual names (such as bicycle, lorry, sports car) testifies to this. If it were not useful to think of all these objects as members of one class, then it is unlikely that the general label "vehicle" would have arisen.

In some respects, therefore, categorisation and classification of stimuli may be seen as crucial in maintaining normal life. If it is accepted that this is the case, then it becomes of great interest to ask exactly how classification systems arise. Are they arbitrary, in the sense that they may vary from person to person; or are they a reflection of real divisions within the actual world of objects? Rosch (1978).

suggests that, "...human categorization should not be considered the arbitrary product of historical accident or of whimsy but rather the result of psychological principles of categorization, which are subject to investigation." (p. 27). To illustrate her point she gives an example of a purely whimsical classification system quoted by Borges (1966). It is a taxonomy of the animal kingdom and has been attributed to an ancient Chinese encyclopedia entitled the Celestial Emporium of Benevolent Knowledge, and as she points out, one of the most interesting things about it is that conceptually it does not exist. It simply does not function in the way that classification systems normally do, and by considering why it does not serve the same functions as other classification systems it becomes possible to discover what some of the principles are that underlie human categorisation. The passage is reproduced below:

On those remote pages it is written that animals are divided into (a) those that belong to the Emperor, (b) embalmed ones, (c) those that are trained, (d) suckling pigs, (e) mermaids, (f) fabulous ones, (g) stray dogs, (h) those that are included in this classification, (i) those that tremble as if they were mad, (j) innumerable ones, (k) those drawn with a very fine camel's hair brush, (l) others, (m) those that have just broken a flower vase, (n) those that resemble flies from a distance. (Borges, 1966, p. 108).

Poetically appealing as this classification system is, it

is not one which has very many practical or linguistic applications: indeed, its very impracticability and illogicality are probably the factors giving rise to its appeal. There are two main respects in which it differs from, for instance, a modern account of the taxonomy of the animal kingdom.

Firstly, not all the categories listed are on the same conceptual level of abstraction, and where differences between levels exist they are not made clear. Many of the categories include several of the others, some include none of the others. For example, category (a), which includes all animals belonging to the Emperor, could conceivably include every other category, because in none of them is it specified that its members do not belong to the Emperor. On the other hand, category (d) (suckling pigs) is very highly specified and does not strictly include members from any other category, although it could be argued that there is some overlap between category (d) and categories (a), (b), (c), (h), (i), (k), (m), and (n). The main point to be made here is that in many cases stimuli which can be legitimately included in one category are not equivalent only to other members of that category but also share many properties in common with stimuli in other categories. Hence the usefulness of placing a given stimulus in a given category is suspect.

Secondly, the differences between categories do not reflect differences within the structure of the actual animal kingdom, at least as we perceive it today. Contemporary classifications, whether at the level of an expert taxonomist or of a young child learning the names of animals, are more likely to be based upon characteristics such as morphological structure and diet, than



they are to be based upon inconstant features such as whether the animals have just broken a flower vase or whether or not somebody has drawn them with a "very fine camel's hair brush".

These two distinctions between the ancient Chinese taxonomy and a contemporary approach reflect two principles which Rosch (1978) proposes and which she suggests underlie human categorisation systems. The first of these she describes as the principle of cognitive economy. She argues that the purpose of categorisation is to enable an organism to gain information from its environment with the least possible effort: categorisation achieves this because, if an object is perceived as being a member of a category, the organism will be able to predict what properties that object possesses (those characteristic of other objects in the category) and also what properties it does not possess (namely those characteristic of objects in other categories). As she points out:

On the one hand, it would appear to be to the organism's advantage to have as many properties as possible predictable from knowing any one property, a principle that would lead to formation of large numbers of categories with as fine discriminations between categories as possible. On the other hand, one purpose of categorization is to reduce the infinite differences among stimuli to behaviourally and cognitively usable pro-

portions. It is to the organism's advantage not to differentiate one stimulus from others when that differentiation is irrelevant to the purposes at hand.

(p. 29).

The second principle of categorisation asserts that the world is perceived as structured information, in that combinations of perceived attributes of objects do not occur randomly or uniformly, but that certain attributes are more likely to appear in combination than others. As Rosch says, "...it is an empirical fact provided by the perceived world that wings co-occur with feathers more than with fur." (p. 29). In order to be useful, in the sense that they save cognitive effort, categories should make distinctions between classes of objects which are actually perceived as being different, on the basis of the extent to which specific combinations of attributes appear. Categorisations should thus reflect the correlational structure of the natural world.

Before discussing the role of categorisation in music perception it is important to mention one further characteristic of categorisation systems, namely that they may be thought of as having both a vertical and a horizontal dimension. The vertical dimension concerns the level of inclusiveness of a category: for instance the term "relatives" denotes a more inclusive category than the term "grandparents" which in turn is more inclusive than "grandmother" which in its turn is more inclusive than "maternal grandmother". Continuing with the same example, the horizontal dimension concerns the segmentation of categories at

the same level of inclusiveness: it is the dimension along which the terms "mother", "father", "aunt", "uncle" vary. As was made clear from examination of the ancient Chinese taxonomy, not all possible levels of categorisation are equally useful: category (1), for example, ("others") is highly inclusive and not at all informative. Rosch (1978) argues that "...the most basic level of categorization will be the most inclusive (abstract) level at which the categories can mirror the structure of attributes perceived in the world." (p. 30). Any categories which are more inclusive than basic level ones are termed superordinate, those which are less inclusive are termed subordinate. Most of Rosch's research has involved an investigation into the characteristics of "basic" objects, or those at the most inclusive level at which there are attributes common to all the members of a given category.

Having discussed the question of why, and how, people categorise the stimuli they encounter, the extent to which the same arguments may apply in the field of music perception will now be considered. ("Perception" here is to be understood as the process of hearing, assimilating, and responding to music.) Very little, if any, research has centred upon this area: although several investigators draw distinctions between different styles or categories of music in terms of the responses they evoke (e.g. Payne, 1967; Bradley, 1971; Geringer & McManus, 1979; Schuckert & McDonald, 1968), they do not actually address themselves to the issues of how and why such stylistic distinctions have arisen, and whether they are actually valid and useful distinctions to make with respect to

their day-to-day occurrence in the real world. Gardner's (1973) study, discussed in Part 2, is something of an exception to the extent that in it he examines the development of the perception of differences and similarities in music. Hargreaves and Colman (1981) also discuss the development of the use of categorical responses to 18 excerpts of music which were taken from compositions of various styles. They argue that responding to a piece of music in terms of a stylistic or categorical label such as "Folk" or "Baroque" or "New Romantic" requires some degree of sophistication and is unlikely to be observed in young children, but point out that "These are preliminary results in a relatively unexplored field..." (p. 19).

It therefore remains an open question as to whether categorisation in music shares all the features of categorisation in other areas. One important notion to be borne in mind is that "music" itself could be considered to be a basic level category, because most of the categories below this level contain many attributes that overlap with other subordinate categories (for instance, "Folk", "Baroque" and "New Romantic" music differ, but all share the attributes of rhythm, tempo, melody and harmony). This argument is of necessity tentative because music can hardly be considered an object in the sense in which Rosch (1978) discusses basic objects, but there is some evidence that the principles of categorisation may be useful in understanding entities other than concrete objects: Rosch found that the same principles appear to apply to "...the cutting up of the continuity of experience into the discrete bounded temporal units that we call events." (p. 43). If it is



therefore provisionally accepted that music itself is a basic category, and also that within the field of music, "musical style" can be considered as synonymous with "category", it follows that any discussion of categorisation or stylistic distinctions within music will relate to subordinate category boundaries rather than basic category boundaries. The implications of this, according to Rosch's discussion, would be:

1) The perceived attributes of "music" as a whole should not be outnumbered dramatically by the perceived attributes of individual styles of music: specifically, if subjects were asked to list the attributes of music and also to list the attributes of 18th Century classical music or Tamla Motown, the number of attributes generated in either of the latter cases should not greatly exceed the number generated in the former case (Rosch, 1978, pp. 32-33).

2) As already mentioned, any given category of music should contain many attributes which overlap with the attributes of other categories of music (Rosch, p. 31).

3) Rosch points out that the most useful name for any item would be expected to be the basic level name. In the present context, this would imply that the term "music" should be used more often and more extensively than subordinate category names such as "expressionist" or "avant-garde jazz".

4) Rosch also argues that, developmentally, categorisation at the level of basic objects should be the first categorisation

used by children, who would only learn subordinate and superordinate category names with increasing age and experience. In the field of music, this might be interpreted as a suggestion that on being played a composition and asked what it is, a child is more likely to simply reply "music" than to say, for example "Boogie-woogie" or "an early English lute song". Ability to respond by referring to specific subordinate categories should develop with age.

All these implications are experimentally testable, but have apparently not yet been investigated. Although they are not tested directly in this thesis, they have been outlined to provide a conceptual framework for Experiment 6 which is described in the next section, and because they may be a useful opening to a systematic examination of the role of categorisation in music perception. All the above implications are based on the premise that music is a basic level category, but it is quite possible that stylistic distinctions between pieces of music are not directly analogous to distinctions between subordinate level objects, and that such an argument is facile in view of the complexity of the area. Factors such as age and musical experience could be influential in determining exactly which "level" is considered to be basic: for instance, what for most people is a subordinate category may come to operate as a basic object category for a specialist.

A further issue, which may be discussed independently of whether or not musical styles can be considered to be subordinate, relates to the way in which one musical style is perceived as being distinct from another. Tversky and Gati

(1978) and Rosch (1978) all stress that stimuli are perceived as structured information rather than as an unpredictable array of attributes, and that humans form categories which reflect these structured perceptions. As Tversky and Gati write:

When faced with a set of stimuli, people often organize them in clusters to reduce information load and facilitate further processing. Clusters are typically selected in order to maximize the similarity of objects within the cluster and the dissimilarity of objects from different clusters. (p. 91).

The implication here is that whether the stimuli are breeds of dogs, subjects discussed in a conversation, or pieces of music, it is people's perceptions of these stimuli which determine where category boundaries should be drawn. Reed (1981), in a commentary to Ghiselin (1981), pointed out that it is very important that the observer's perceptions of category boundaries should actually correspond to some real disjunction between stimuli, and says that Rosch "...often emphasizes categorical organization as a 'contribution of the processor' (Rosch & Lloyd, 1978, p. 2) to the world. Obviously, such contributions will not be adaptive, except where the contributions are faithful to the way the world is." (p. 298). This argument is relatively clear-cut so far as concrete and functionally defined objects (such as articles of furniture) are concerned. It is in fact clear-cut when applied to any category or group of categories for which there is recourse to an objective method of

distinguishing between members: for instance, different species of animal may be distinguished on the basis of whether they have fur or feathers, or whether they are oviparous or give birth to live young. However, in the realm of music or any other stimuli which are primarily aesthetic, it is conceivable that distinctions between categories are made on a more subjective basis, and therefore do not correspond so directly to "real-world", objective differences. To take an example, Chapman and Williams (1976) were able to present the same piece of music to two groups of adolescents and describe it in one case as "progressive pop" and in the other as the work of a contemporary "serious" composer without arousing any suspicion in their subjects that the piece was not actually representative of the style given. This hypothesis (namely that distinctions between styles of music are made upon a more subjective basis than distinctions between [e.g.] varieties of plants or types of car engines) is purely speculative, but it is worthy of consideration especially in the light of Chapman and Williams' findings, one of which was that subjects to whom music was described as "progressive pop" were significantly more favourably disposed towards it than were those to whom it was described as "serious": despite the fact that both groups of subjects were judging the same music. One interpretation of this might be that, for some people at least, the category to which a piece of music is seen to belong can be an important determinant of their liking for that music. Paradoxically, however, the grounds upon which they distinguish categories may be equivocal. Alternatively, Chapman and Williams' findings



might be a function of the particular composition they used: further investigations which involve more than one composition would throw light upon this.

Many tentative suggestions have arisen in this section about the role of categorisation in human beings' responses to music. Because of the dearth of empirical evidence it is difficult to predict which of these suggestions might be an accurate or useful description of the processes involved. However, it could be argued that distinctions between styles of music and between the accompanying verbal labels for these styles are extremely important, if only from the point of view of being able to talk about musical taste, or describe compositions or composers to people who are unfamiliar with them. This area might prove to be a fruitful one for researchers because it underlies many different aspects of music appreciation such as the use of language in responses, the development of specific preferences, and the importance of peer group norms in musical taste. Experiment 6 investigates some aspects of the way in which the use of stylistic labels develops with age.

#### 4.2: The Present Research

Experiment 6 constitutes a very exploratory investigation of the way in which the categorisation of music changes with age. The results of Experiments 1 and 2 and of Hargreaves and Colman's (1981) research suggested that the tendency to describe music by using stylistic labels increases with age. Machotka (1963) and Gardner and Gardner (1970) also argued that the concept of an artistic or musical style emerges only after the age of 11 or 12 years, because preadolescent children neither speak in terms of style nor spontaneously classify works on that basis. Bearing these findings in mind, 11- and 12-year-olds were chosen to be the youngest subjects in the present study, because it was felt that this was the earliest age at which children might be expected to have an understanding of the use of stylistic labels. The second age group used comprised 14- to 15-year-olds, and the third group consisted of adults of 18 years and over. It was felt that the use of labels by the 14- to 15-year-olds might be of particular interest in view of the importance of music to this age group (Johnstone and Katz, 1957; Chapman and Williams, 1976).

All subjects were asked to listen to 12 unidentified excerpts of music, about which no information was provided. All 12 excerpts had been selected so that they represented a wide range of styles but at the same time were difficult to

classify stylistically. After hearing each excerpt, subjects were required to indicate which stylistic labels from a list of 19 they thought might aptly be used to describe it: they were encouraged to describe each excerpt by as many of these labels as possible. There were three main areas of investigation in the present experiment. First, on the basis of Machotka's (1963, 1966) and Gardner and Gardner's (1970) research, it was predicted that the older subjects in the present study might find it easier to classify the excerpts than would the younger subjects. The former group might therefore be less inclined than the latter to use a large number of different stylistic labels to describe each excerpt because there might be a higher level of agreement among subjects as to which labels were appropriate. Alternatively, bearing in mind that the excerpts were selected on the basis of their unfamiliarity and their ambiguous stylistic qualities, it is possible that older subjects might have a rather more flexible approach to the application of stylistic labels and would use more, different labels than would younger subjects. Thus the first issue under investigation in the present study was whether there were any developmental trends in the number of labels applied to each excerpt by subjects.

Bearing in mind the nature of the labels (see "Method" section for a comprehensive list) it was predicted that subjects would use some labels more often than others; because some might be thought appropriate for only one or two, or even none, of the excerpts heard. By counting the number of excerpts described by each label, it was hoped to obtain a

measure of the way subjects perceived the inclusiveness of a given label. For instance, labels such as "Classical" or "Avant Garde" might be found to be more inclusive than rather specific labels such as "Brass Band" or "Hymn", especially in view of the stylistically ambiguous characteristics of the stimulus music. Thus the second issue under investigation was a comparison among the three age groups of the perceived inclusiveness of each label. Other than there being no systematic age-related changes, it was felt that there might be two possible outcomes of such a comparison. Firstly, with increasing age, subjects might overtly acknowledge that any one stylistic label might encompass a wider range of music than they previously realised. This would result in any one label being used to describe more excerpts by the older than by the younger subjects: in other words, each label could be seen as more inclusive with increasing age. Secondly, it is possible that as subjects become more familiar with the general concept of musical style, the defining parameters of any one style become sharper and more rigid; therefore, as they get older, subjects might tend to use any given label to describe a progressively smaller number of excerpts. This would constitute a decrease in the perceived inclusiveness of each label with increasing age.

The issue just described relates to the way in which the perceived inclusiveness of each label alters with age, and the third and last issue under investigation took this a step further. It involved an examination of the extent to which a label that was used often by one age group was likely to be



used often by the other two age groups: it is possible that each age group perceives a different label as being the most inclusive. A hypothetical example of this might be that adolescent subjects, with their detailed knowledge of contemporary music, might not use the "Punk/New Wave" label at all, because none of the stimulus material actually belonged to this category; whereas adult subjects who are not necessarily as "in touch" with current musical developments might use this label in a less discriminating way, applying it to any of the stimulus music which sounded unusual to them.

In summary, the experiment was designed to investigate the following: (a) Whether the number of stylistic labels applied by subjects to each musical excerpt alters with age, (b) whether the number of excerpts described by each label alters with age, and (c) whether a label which is thought of by one age group as being very inclusive is likely to be thought of in the same way by either of the other age groups. In addition, the results of the present study were to be used as a basis upon which to select four excerpts of music ambiguous in one particular respect, namely that subjects were likely to describe them as both "Rock Jazz" or "Modern Jazz" and "Classical" or "Modern Classical"; for use in Experiment 7.

### 4.3: Experiment 6

#### The Application of Stylistic Labels to Ambiguous Excerpts of Music

##### Method

##### Subjects.

Sixty subjects took part in the experiment, 20 aged 11 to 12 years, 20 aged 14 to 15 years and 20 aged 18 and over. (Ten members of the latter group were over 30.) Within each age group there were approximately equal numbers of males and females. The two younger groups were tested in schools in Leicestershire, 10 at a time, and the adults were tested in smaller groups numbering one to eight in the Psychology Department at Leicester University.

##### Musical material and questionnaires.

One excerpt was selected from each of the 12 compositions listed below. Criteria for the selection of compositions were that a wide range of instrumentation should be represented, that the compositions should be relatively little known, and that they should not be of any single readily identifiable musical style. The last two criteria were checked by examining the first 10 response sheets from adult subjects, and it was found that only one excerpt (from piece no.10) was familiar to

one subject, and that a very wide range of stylistic labels had been applied by all subjects.

Each excerpt was 40 seconds long. The excerpts were recorded on a cassette tape in the following order: (a) "Space Flight" by Ornette Coleman, (b) "Forms and Sounds" by Ornette Coleman, (c) "Ano Zero" by Egberto Gismonte, (d) "An American in England" by Pierre Moerlan, (e) "A Leicester Jazz Suite" by D. Hargreaves, (f) "Chamber Concerto for 13 instrumentalists" by Ligeti, (g) "Ard na Greine" by Pierre Moerlan, (h) "Salvador" by Egberto Gismonte, (i) "Entr'acte Vergine Bella II" by Charles Wuorinen, (j) "Malvern Hills" by Pete Cooke, (k) "A Flock Descends into the Pentagonal Gardens" by Toru Takemitsu, (l) "Piva" from "Italian Airs and Dances" arranged by J. Dalza.

Instructions on each response sheet were as follows:

Please listen to these extracts of music. After hearing each one, place a tick next to any name that you think might describe the type of music it is. For instance, if you feel that the first piece might be "Brass Band", "Trad Jazz" or possibly "Classical", then you should tick all three. It is important that you tick every category name which you feel could describe the extract, even if you are very unsure about it; there is no limit to the number of ticks you are allowed.

Below the instructions was a 19 x 12 grid. The 19 rows were for 19 stylistic labels, and there was one column for

each of the 12 excerpts to which subjects were required to listen. A sample response sheet appears in Appendix 4.3.1. The stylistic labels were intended to cover as broad a spectrum of musical styles as possible, and were derived from a study by Hargreaves and Colman (1981) with this in mind. The extent to which this was actually achieved for all age groups will be discussed later. The 19 stylistic labels were: Brass Band, Country and Western, Trad Jazz, Modern Classical, Punk/New Wave, Modern Jazz, Modern Folk, Rock Jazz, Opera, Trad Folk, Avant Garde, Muzak, Hymn, Classical, Romantic, Blues, "Musical", Non-European, and Soul.

#### Procedure.

Subjects were presented with one response sheet each and were asked to read the instructions carefully. After ensuring, by further explanation if necessary, that all subjects understood the task, the excerpts were played one by one, stopping the tape at the end of each excerpt in order to give all subjects enough time to mark their response sheets. It was stressed that there were no correct or incorrect responses and that for any one excerpt there was no limit to the number of stylistic labels a subject could mark as being appropriate. Subjects were asked to circle the number referring to any piece with which they were familiar and to name it if possible. When all 12 excerpts had been heard, subjects were asked to make a note of their age and sex on their response sheets, which were then collected.



## Analysis of Responses and Results

Two, two-way analyses of variance were performed, each with repeated measures on one factor. The first examined the number of labels applied to each excerpt by each of the three age groups. The main effects for age and for number of labels applied were significant ( $F[2,57] = 4.40$ ;  $p < 0.05$  and  $F[11,627] = 3.72$ ;  $p < 0.001$  respectively), as was the interaction effect ( $F[22,627] = 3.20$ ,  $p < 0.001$ ). The results of this analysis are summarised in Table 4.3.1. The second analysis examined the number of excerpts described by each label, for each of the three age groups. Again, all three F-ratios were statistically significant: for age,  $F(2,57) = 3.77$ ,  $p < 0.05$ ; and for the number of excerpts described by each label,  $F(18,1026) = 16.20$ ,  $p < 0.001$ . For the two-way interaction  $F(36,1026) = 7.04$ ,  $p < 0.001$ . The results of this analysis are summarised in Table 4.3.2. Main effects and interactions for the two analyses are illustrated graphically in Figures 4.3.1, 4.3.2 and 4.3.3 (first analysis) and 4.3.4, 4.3.5, and 4.3.6 (second analysis).

A Pearson's product moment correlation coefficient was computed, using the mean number of excerpts described by each of the 19 labels as raw data, between the 11- to 12- and 14- to 15-year-olds, to establish whether labels perceived as inclusive by one age group were similarly perceived by the other age group. The correlation coefficient was 0.65,  $df$  17: this reached statistical significance ( $p < 0.01$ ). Correlation coefficients were also computed between the same data for the

adults and for the other two age groups, and the resulting values for  $r$  were: adults and 14- to 15-year-olds;  $r = .39$ ,  $df = 17$ ; adults and 11- to 12-year-olds;  $r = .43$ ,  $df = 17$ . Neither of these values were statistically significant at the 0.05 level. The labels which had been used to describe each excerpt were examined in order to find out which excerpts were most often perceived as representative of "Rock Jazz" or "Modern Jazz" and "Classical" or "Modern Classical" styles simultaneously. The four excerpts which were most often perceived as ambiguous in this respect were: "Forms and Sounds" by Ornette Coleman, "A Leicester Jazz Suite" by David Hargreaves, "Chamber Concerto for 13 Instrumentalists" by Ligeti, and "Entr'acte II: Vergine Bella" by Charles Wuorinen.

### Discussion

By examining Figure 4.3.1 it can be seen that the adults and 11- to 12-year-olds apparently used more labels to describe each excerpt than did the 14- to 15-year-olds. Before attempting to interpret this, it must be borne in mind that both the second main effect (Figure 4.3.2) and the interaction (Figure 4.3.3) of this analysis were statistically significant. Figure 4.3.3 suggests that the 11- to 12-year-olds in particular tended to apply a large number of labels to the excerpts appearing in the latter half of the tape, and it is possible therefore that the high mean achieved by this age group was artifactual, being due partly to an order effect. It is interesting that this should occur only in

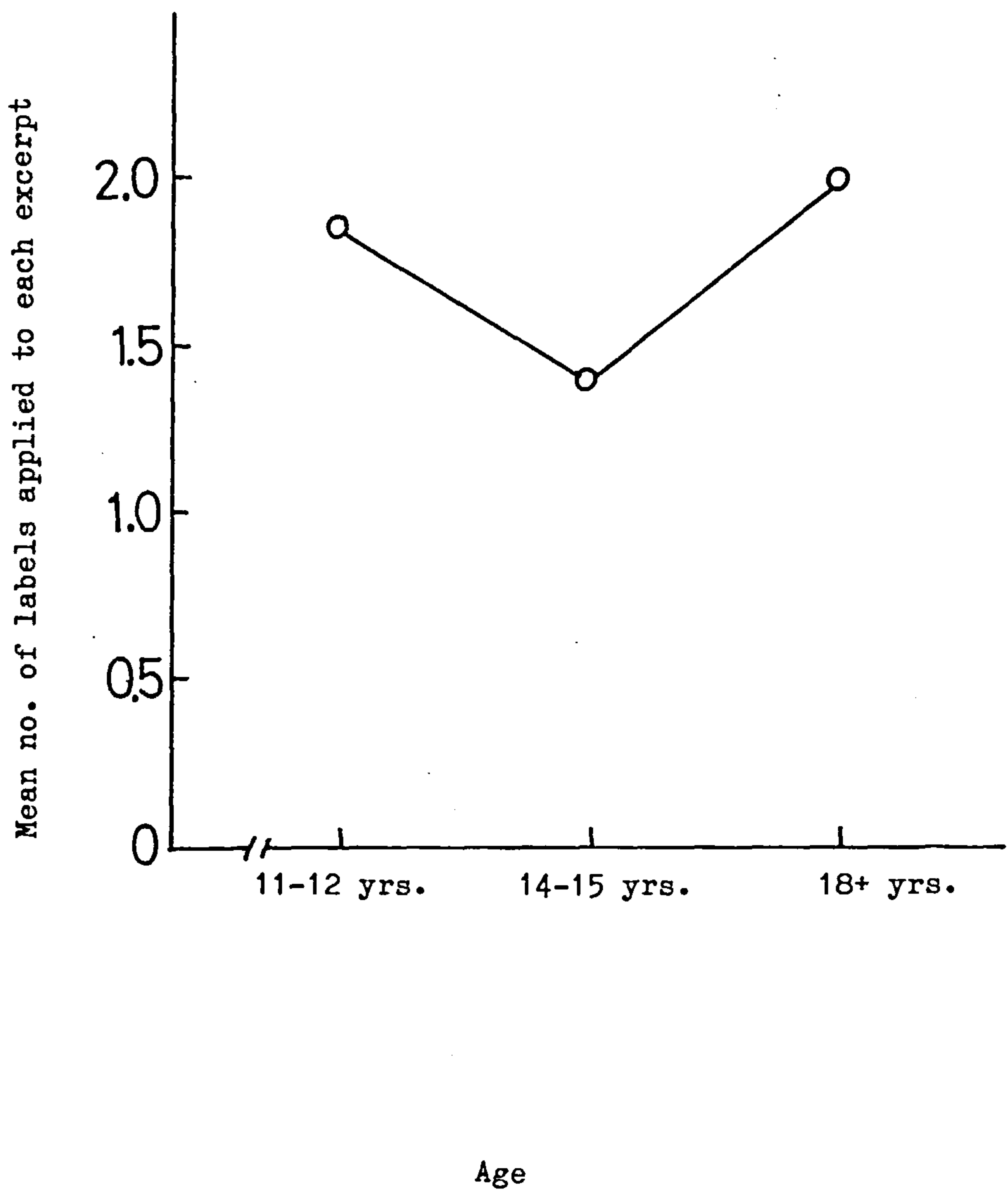


Figure 4.3.1: Significant main effect.  
The effect of age on the mean number of labels applied to each excerpt, per subject.

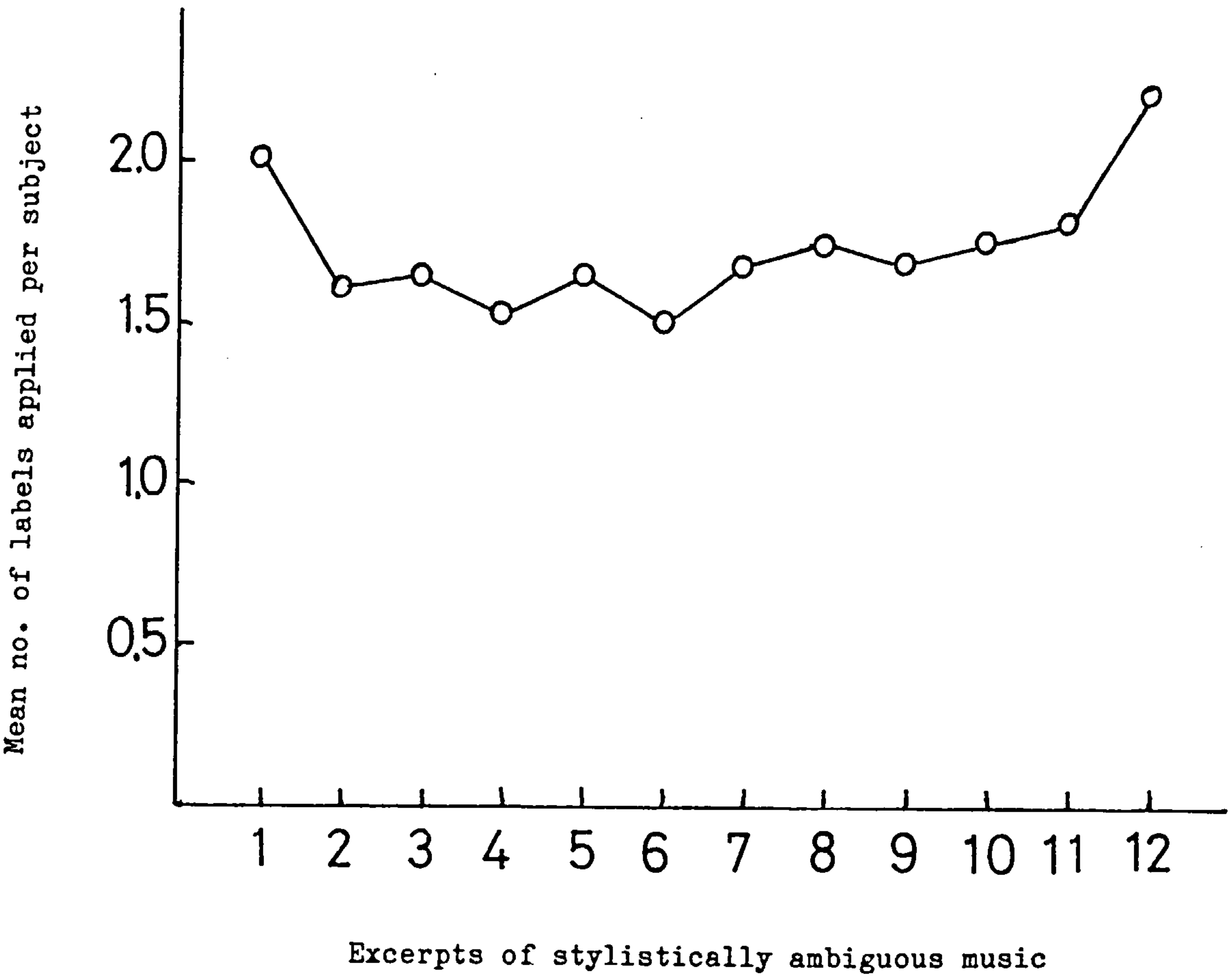


Figure 4.3.2: Significant main effect.  
Mean number of stylistic labels applied to each of 12 excerpts of ambiguous music, per subject. Numbers 1 to 12 give the position of each excerpt on the tape. Details of the excerpts are given in Part 4.3.



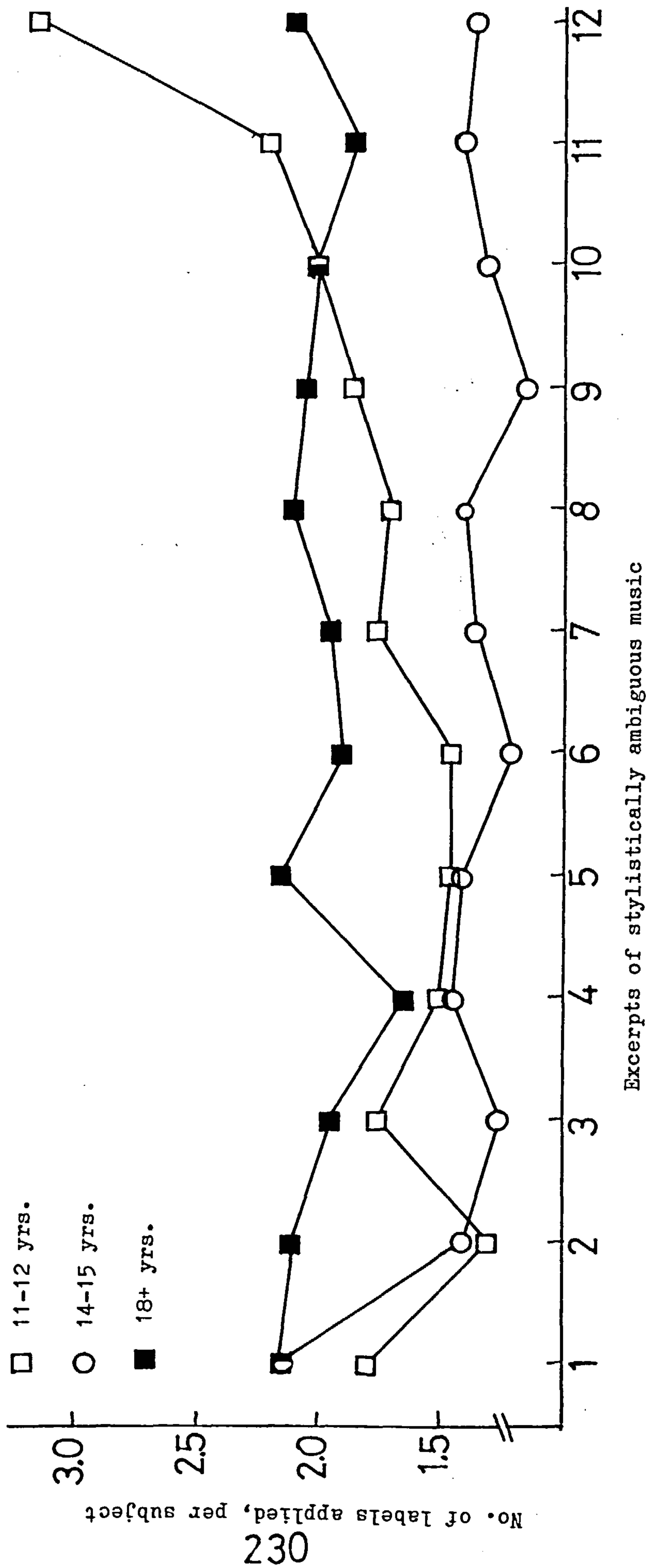


Figure 4.3.3: Significant two-way interaction. Interaction between age and the number of labels applied to each excerpt, per subject. The numbers 1 to 12 on the X-axis give the position of each excerpt on the tape. Details of the excerpts are given in Part 4.3.

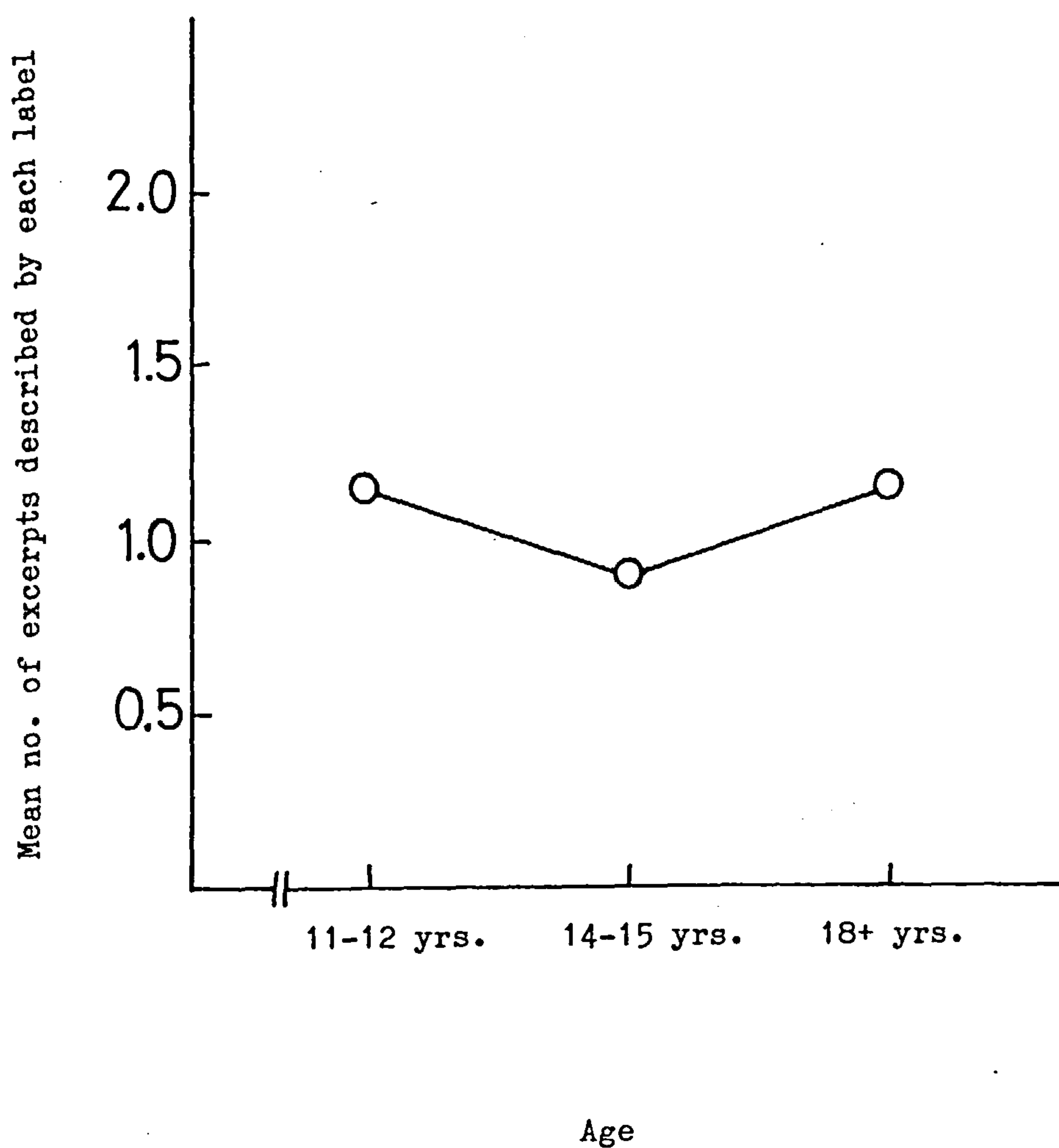


Figure 4.3.4: Significant main effect.  
The effect of age on the mean number of excerpts described by each label, per subject.

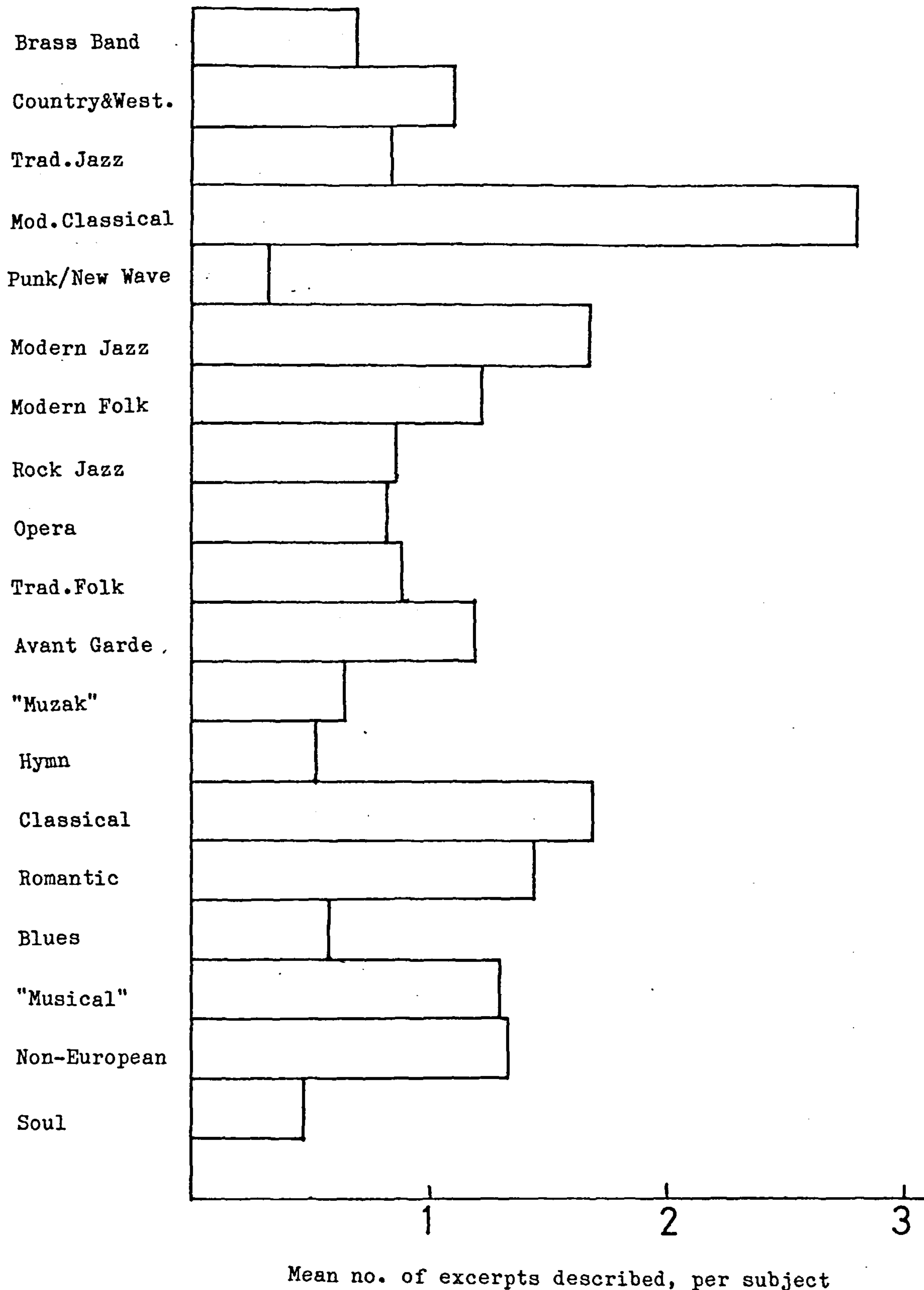


Figure 4.3.5: Significant main effect.  
The number of ambiguous excerpts described by each of 19 stylistic labels, per subject.

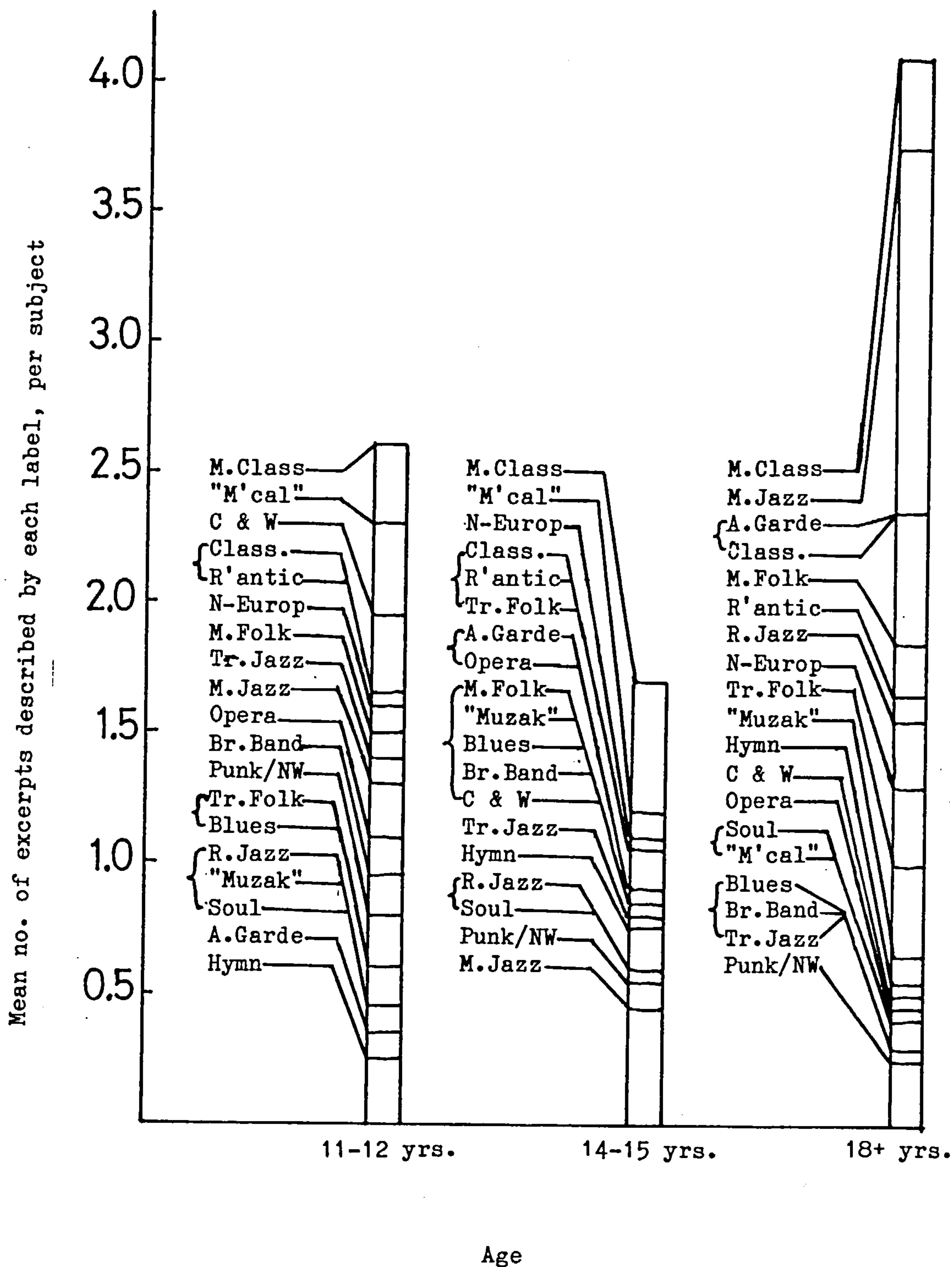


Figure 4.3.6: Significant two-way interaction.  
The interaction between age and the perceived inclusiveness of each of 19 stylistic labels (as measured by the number of excerpts described by each label).



Table 4.3.1

Analysis of variance summary table,  $N=60$   
The effects of age on the number of stylistic labels applied  
to 12 excerpts of music.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	2	22.593	4.40	<0.05
ERROR	57	5.132		
B	11	2.380	3.72	<0.001
AB	22	2.046	3.20	<0.001
ERROR	76	0.568		

A=age, 11-12 years, 14-15 years, and over 18 years  
(independent factor).

B=stylistically ambiguous excerpts (repeated measures  
factor).

Table 4.3.2

Analysis of variance summary table, N=60  
The effects of age on the number of excerpts described by 19  
stylistic labels.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	2	12.334	3.76	<0.05
ERROR	57	3.277		
B	18	19.306	16.20	<0.001
AB	36	8.386	7.04	<0.001
ERROR	1026	1.192		

A=age, 11-12 years, 14-15 years, and over 18 years  
(independent factor).  
B=stylistic labels describing excerpts (repeated measures  
factor).

the youngest age group, but is perhaps explicable in terms of their greater susceptibility to fatigue. It is also possible that some sort of order effect was at work at the beginning of the tape, where the two younger age groups applied noticeably more labels to the first than to subsequent excerpts. However, overall, results suggested that any developmental trend here was not simply a question of a steady increase or decrease with age in the number of labels applied to each excerpt, because the middle age group were apparently the least willing to apply a large number of labels overall. To some extent therefore, the prediction based on Gardner's (1970) suggestion that stylistic sensitivity only emerges fully after the age of 11 to 12 years has been upheld: it was argued in Part 4.2 that if this were the case, subjects would find it easier to classify excerpts as belonging to one style or another, and would therefore apply fewer labels, as they grew older. However, this is apparently true only up to a certain age because adult subjects applied more labels to each excerpt than did either of the two other age groups. Before going on to discuss the extent to which this effect might be partially due to an alteration with age in the nature of the labels selected, it is worth mentioning that the 11- to 12-year-olds were still receiving formal music teaching at school, the data being collected during what was normally a music lesson. The adult(s) involved in teaching them did not presumably differ significantly from those who generated the initial list of 19 stylistic labels, so the 11- to 12-year-olds might have been more likely to come into contact with the everyday application

of these labels than might the 14- to 15-year-olds, who were no longer receiving any formal music tuition at school. The two younger age groups may not have understood the labels as well as the adults because they were produced by adults; and this could have affected the 14- to 15-year-olds more than the 11- to 12-year-olds because the latter group were regularly in a situation in which they might be exposed to similar labels.

The second analysis of variance (Table 4.3.2) investigated the way in which the perceived inclusiveness of each label altered with age. Two possibilities were proposed in Part 4.2; first, that as the age of subjects increases, the conception of what constitutes a given musical style becomes more rigid, resulting in each label becoming less inclusive. Alternatively, subjects might with increasing age and experience come to the conclusion that it is difficult to precisely define the parameters of any one musical style. In the context of the present experiment, this would result in the older subjects applying each label to more excerpts than did the younger subjects. Figure 4.3.5 shows that in fact the adults and the 11- to 12-year-olds applied each label to more excerpts than did the 14- to 15-year-olds: again demonstrating an apparent discontinuity in a developmental trend.

Certain labels were applied more often than others over all age groups, as can be seen from Figure 4.3.5. "Modern Classical" was the most commonly applied label, being applied to a mean of 2.8 excerpts, followed by "Modern Jazz" and "Classical", "Romantic", "Modern Folk", "Non-European", "Musical", "Avant-Garde", "Country and Western", "Trad Folk"



and "Rock Jazz", "Trad Jazz" and "Opera", "Brass Band", "Muzak", "Blues", "Punk/New Wave", "Hymn" and "Soul". This order of perceived inclusiveness did not hold for all age groups, as can be seen from Figure 4.3.6. All three age groups applied the "Modern Classical" label to more excerpts than they did any other label, but the next most inclusive label differed among age groups; for the two younger groups it was "Musical" whereas the adults used "Modern Jazz". It therefore seems likely that the two younger groups misunderstood the sense in which "Musical" was meant: it was intended to refer to the style of music which appears in film and stage "musicals" such as "My Fair Lady", but it is conceivable that these subjects understood it to mean "musical" in an adjectival sense; as the opposite of "unmusical". It should also be pointed out that, because the list of stylistic labels was generated by adults, several years ago (1974), current styles of popular music (e.g. New Romantics) were not represented at all, and labels that might be of little meaning to adolescent and younger listeners were in some respects over-represented, as exemplified by the distinction between three types of jazz, or the inclusion of "Avant-Garde". The finding that adults used more labels to describe each excerpt than did the younger groups (Figures 4.3.1 and 4.3.3), as well as the finding that they perceived the more general labels (such as "Modern Classical", "Modern Jazz", "Avant-Garde" and "Classical") as being appropriate to a larger number of excerpts than did the other age groups (Figure 4.3.6), could be a reflection of the age bias in the initial method of

selection of the labels.

It is interesting that an increase in the number of excerpts described by each label did not occur uniformly with increasing age, suggesting that between the ages of 11 to 12 and 14 to 15 subjects became less willing to use the same label to describe many different excerpts. This implies an increasingly conservative approach to the use of the 19 labels up to the age of 14 or 15. At some time after this age, subjects become more flexible in their application of stylistic labels and readily use the same label to describe several excerpts (Figure 4.3.6).

The correlation between the 11- to 12-year-olds' and 14- to 15-year-olds' rankings of the 19 labels according to their perceived inclusiveness reached statistical significance, whereas those between the other pairs of age groups did not. The implication of this is that any label which is perceived as more (or less) inclusive by the 11- to 12-year-olds was likely to be perceived in the same way by the 14- to 15-year-olds, and vice versa, whereas no such relationship held between the responses of the adults and those of either younger group. This again could be a reflection of the fact that both the 11- to 12-year-olds and the 14- to 15-year-olds were attempting to apply stylistic labels which were generated by an age group other than their own.

As has become apparent during the discussion, there are some respects in which the design of this experiment might be improved if it were to be repeated. For instance, several different random orders of presentation could have been used

to help rule out any possible order effects. In addition, the stylistic labels supplied may not have been appropriate for all age groups equally, because they were generated by adults, and the range of excerpts from which the excerpts of music were selected may have been too restricted: the stylistic labels were selected because they had been found to cover a wide range of musical styles (Hargreaves and Colman, 1981), but there were no such external criteria for selection of the excerpts. Nevertheless, the findings are suggestive in several respects which are discussed in Part 4.4.



#### 4.4: Conclusions (Part 4)

The following conclusions can be drawn: the 14- to 15-year-old subjects on the whole applied fewer labels to each excerpt than did the 11- to 12-year-olds (Figure 4.3.1), and in many cases individual 14- to 15-year-old subjects provided only one label for each excerpt. This suggests that they did not readily think of an excerpt as possibly belonging to two styles at once. Inspection of the response sheets indicated that there was little agreement among subjects as to which style(s) actually were represented by each excerpt, so the 14- to 15-year-olds might be said to exhibit a high "among-subjects", but a low "within-subjects", tolerance of stylistic ambiguity. This conservative approach was also reflected in the results of the second analysis of variance (Figure 4.3.4) which suggests that they perceived the labels as being less inclusive than did the other age groups. The trend towards "conservatism" between the ages of 11 and 15 years was reversed at some time later, because adult subjects perceived the labels as being more inclusive. This could suggest that with increasing maturity subjects realise that there are not necessarily any clear-cut distinctions between one musical style and another. Adults also applied more labels to each excerpt than did other age groups (Figure 4.3.2).

It should be pointed out that all subjects, including the adult group, found it very difficult to complete the task set in this experiment. This may have been because the stylistic



labels provided were not seen as appropriate, although they had been selected because there was evidence to show that they covered a wide range of stylistic possibilities. In addition, it may be that the verbal labelling of different styles of music is not important to many people, except with reference to any styles with which they are particularly familiar: and most musical excerpts in the present experiment were not familiar to the subjects. Developmentally there appeared to be a trend from the age of 11 or 12 to use stylistic labels with progressively more caution: having selected a particular label as appropriate, 14- to 15-year-olds were unwilling to admit that other labels might also be appropriate. This finding might imply that the 14- to 15-year-olds were attempting to be unrealistically precise in defining any one musical style; and this was reflected in the low level of agreement among these subjects as to which label was appropriate for each excerpt. The adult group, however, were prepared to acknowledge (at least covertly) that the concept of musical style is somewhat ill-defined. They showed a tendency to use several stylistic labels to describe any one excerpt, and there was also a relatively high level of agreement among subjects as to which labels were appropriate for each excerpt. This high level of agreement would result in a greater understanding among adults discussing styles of music than would result from the low level of agreement shown by the adolescent subjects. This suggestion concurs with Rosch's (1978) hypothesis that subordinate category names are learned and used later than basic level names, and supports the idea that musical category

names do not correspond to basic level names. Indeed, the comments subjects made about the difficulty of the task suggest that even among adults the use of stylistic labels may not be consistent, and that more information may be needed in order to make a category judgement than can be obtained by listening to a short excerpt of music without any social or other context.

One of the most interesting findings to emerge concerns the responses of the 14- to 15-year-olds, which did not fit into any continuous model of developmental progress, but rather suggested that a U-shaped trend may be more appropriate. This age group is investigated further in the next experiment.

PART 5 : THE EFFECTS OF EXTRA-MUSICAL INFORMATION ON  
RESPONSES TO MUSIC

5.1 : Some Previous Research

Experiments reported so far in this thesis have investigated the effects of variables intrinsic to music, such as tempo and melodic complexity, on listeners' responses. In Experiment 5 extra-musical variables were also manipulated in that some subjects were asked to imagine that they were angry, and the resulting effects on their musical preferences were investigated. An examination has also been made of the nebulous concept of musical style; however, the difficulty of defining a musical style in terms of musical variables alone was pointed out in the discussion of Experiment 2 and in Part 4.1. It is possible that the factors which will be taken into account when deciding to what style a piece of music belongs amount to more than, for example, an evaluation of its instrumentation, tempo, and harmonic, melodic, and rhythmic characteristics alone. The results of Experiment 6 tend to support this suggestion in that subjects were by no means unanimous in assigning stylistic labels among a selection of excerpts to which they listened. This suggests that other determining factors may be worthy of consideration, such as the extent of previous musical experience, the context within which the music is heard, or additional information received from a music teacher or a



friend. This section will examine the effects of extra-musical information on responses to music. The hypotheses under investigation have particular relevance for the assessment of the importance of social factors in music perception.

Chapman and Williams' (1976) research offers confirmation of the idea that external information can be important in determining how people respond to music. Following on from Asch (1948), who argued that people's perceptions of a piece of artwork, and hence their aesthetic reactions, may change as a function of the context within which the work is perceived and of the nature of available information relevant to the work, they took an excerpt of music and ascribed to it a different status for each of three groups of subjects by systematically varying experimental instructions. The subjects involved were adolescent schoolchildren who were progressive pop enthusiasts and who were in general negatively disposed towards "serious" music. The "high status music" group was told that the music they were about to hear was a piece of "progressive pop" music, the "low status music" group was told that it was a piece of "modern serious" music, and the control group was given no information about the music. Factor analysis of ratings given by the subjects revealed that those subjects in the "high status music" group were more favourably disposed towards the excerpt than were those who were told that it was modern serious music. Subjects in the control group, who had been given no information and were hence unable to draw on social norms with confidence, were unsure about their reactions to the music.

Rigg (1948) also manipulated the perceived prestige of a



piece of music, in this case by Wagner, by playing it to three groups of American college students. This took place just before World War Two. On the first hearing, all three groups of subjects were given no information about the piece, but on the second hearing one group was told that Wagner was identified with Hitler and German Nationalism. This group reported a smaller increase in enjoyment than the group who were given no information on the second hearing. They in turn reported a smaller increase in enjoyment than the third group for whom the music had been described in romantic terms on the second hearing.

Chapman and Williams (1976) initially suggested that, "The importance of social prestige in aesthetic reactions is presumably a direct function of a person's commitment to the art in question and an inverse function of background knowledge that he may have concerning the piece under scrutiny..." (p. 62). The results of their study confirmed the first part of this hypothesis: they selected subjects on the basis of their commitment to progressive popular music, choosing only those who viewed it favourably, and found that in this subject population manipulation of the perceived status of a piece of music did produce changes in subjects' evaluations of it. With respect to the second part of their hypothesis (that the importance of social prestige may be an inverse function of background knowledge a person possesses concerning the particular work of art in question) their results were not so illuminating, because the composition chosen was unfamiliar to the subjects. There is research which suggests that expertise within a particular area

(as opposed to knowledge about the specific work of art under examination) is not inversely related to the importance of social prestige: for instance, Wieck et al. (1973) investigated the reactions of two jazz orchestras to compositions which were described as either serious, commercial, or neutral. In this case, the subjects might safely be assumed to be fairly knowledgeable about jazz compositions in general, although not familiar with the particular compositions presented. Their performance and later recall of a composition were both found to be adversely affected by the information that the composition was non-serious. However, this evidence by no means disconfirms Chapman and Williams' second hypothesis, because expertise in a particular area is presumably a direct function of a person's commitment to that area, and it would therefore be predicted that general expertise and the importance of social prestige might co-vary, up to the point where the level of general expertise becomes so high that a person has much knowledge about specific compositions.

Another experiment which involved manipulation of the perceived prestige of musical compositions was performed by Duerksen (1972). It differs somewhat from those described by Chapman and Williams (1976) and Rigg (1948) in that subjects were asked to rate the performance and technical qualities of what they heard instead of rating subjective and affective characteristics. Duerksen played the same recorded piano performance twice to music and to non-music specialists. The control group was simply required to rate the standard of excellence in each performance, without being told that the

two performances were identical. The experimental group was told that one performance was by an eminent professional pianist, Wilhelm Backhaus, and that the other recording was taken from an audition tape submitted by a student seeking admission to a graduate music course. Order effects were controlled. Duerksen found that both music and non-music majors gave lower ratings when they had been told the performance was by a student than they did when told it was by a professional pianist. These findings are interesting because even those subjects specialising in music were influenced by information about the prestige of the performer: and to this extent they offer corroboration of Chapman and Williams' results where subjects with a high degree of commitment to a particular musical genre were susceptible to alterations in the perceived status of music within that genre. A second noteworthy aspect of Duerksen's research is, as he points out, that, "The relatively objective characteristics, such as pitch and rhythm accuracy, did not seem to be any more or less influenced than the relatively subjective characteristics, such as interpretation and overall effect." (p. 271).

All the studies mentioned so far have either used subject populations which displayed a relatively high level of commitment to the particular style of music used as a stimulus (with the exception of Duerksen's non-music majors) or have involved providing fairly emotive information about the music (as in Rigg's 1948 research). In other words, in all these cases, it has clearly been the perceived status or prestige associated with the music which was the independent variable.



It should also be pointed out that there exists some research, cited by Chapman and Williams (1976), which has failed to substantiate the importance of social prestige in aesthetic judgements: for instance an experiment using poetry by Michael et al. (1949) and one using extracts from plays as the aesthetic stimuli (Frances, 1963).

A question remaining, therefore, is whether aesthetic judgements may be affected by information which is less obviously emotive than that given by Rigg (1948), using a subject population which has not been selected for its commitment to one particular musical genre.

Specifically, it is possible that subjects might judge music differently according to which of two styles they have been told it represents, even although the subjects have not been chosen on the basis of their liking for one of those styles and their dislike for the other; in other words it is possible that extra-musical information may affect responses to music even when that information does not alter the perceived prestige of the music in a systematic way. A second question remaining concerns the extent to which the effects of external information on responses to music change with age: none of the research discussed uses subjects under the age of 14, and since children younger than this are also exposed to information about music, whether at school from teachers, or from family and peer group, it would be interesting to find out whether pre-adolescents are affected by this information to the same extent as adolescents. These questions are investigated in Experiment 7, and the specific hypotheses to be tested are formulated in Part 5.2. It



is of course quite probable that these two issues are interdependent to some extent, because information about the style of a piece of music may have the effect of altering its perceived prestige for one age group, but not for another. For instance, rock music might be seen as more prestigious than classical music by adolescents, but younger children might not have this view: and adults might associate more prestige with classical than with rock music.

## 5.2: The Present Research

Experiment 7 is similar to Chapman and Williams' (1976) study in that it involves the presentation of musical stimuli under one of two guises: either as classical or as popular compositions. It differs from Chapman and Williams' research, however, in two important respects. Firstly, Chapman and Williams' research used only one composition as a stimulus, and it is possible that the effects they observed could have been a function of the particular composition they chose. In order to control for this contingency, four excerpts were selected for use in Experiment 7. It had already been established that these excerpts were the most likely, out of the 12 excerpts rated in Experiment 6, to be described as both "Classical" or "Modern Classical" and "Modern Jazz" or "Rock Jazz". This increased the likelihood that subjects would find a presentation of the excerpts as either "contemporary classical" or "modern progressive rock" plausible. Secondly, the range of subjects' ages was extended relative to that in Chapman and Williams' study; there were three age groups of 10 to 11 years, 14 to 15 years and 18 to 19 years respectively. It was hypothesised that the 14- to 15-year-olds might be influenced by stylistic information to a greater extent than either of the other age groups, because Chapman and Williams found that music is an important means of social identification at this age, with 14- to 15-year-olds' preferences being related to sex and educational groupings within their school. Similar findings are

reported by Murdock and Phelps (1972).

It was mentioned in Part 5.1 that Duerksen (1972) found that the perceived status of a performer affected ratings of both objective and more subjective characteristics of the music. Two types of rating were required in the present experiment, one of which was wholly subjective (liking), the other less subjective in the sense that it was evaluative rather than affective (quality). The latter was intended as a measure of the perceived eminence and technical competence of the composer, and is a measure described by Hargreaves et al. (1980) as being correlated with, but different from, affective ratings. The latter researchers found that adult subjects displayed fragmentation of "liking" and "quality" ratings in that they were inclined to give higher liking ratings for popular than for classical music, but were likely to say that classical music was nevertheless of a higher quality than popular music. In the context of the present experiment, it was thought that liking and quality ratings might be differentially affected by information as to the style of the music. It is possible, for instance, that subjects who are told that ambiguous excerpts are "modern progressive rock" might like them better, but give lower quality ratings, than subjects who are told that the excerpts are "contemporary classical" music.

To summarise, Experiment 7 investigated the influence on liking and on quality ratings for stylistically ambiguous excerpts of differential information about the style of those excerpts. The excerpts were described as either "modern progressive rock" or "contemporary classical". A comparison

among three age groups of the extent to which this influence operates was also undertaken: one possible outcome might be that adolescents to whom the excerpts were presented as "contemporary classical" might like them less than did adolescents to whom the excerpts were described as "modern progressive rock".



### 5.3 : Experiment 7

#### The Effects of Stylistic Context on Preference and Quality

##### Ratings of Ambiguous Excerpts of Music

#### Method

##### Subjects.

There were 40 subjects in each of three groups aged 10 to 11 years, 14 to 15 years and 18 to 19 years respectively. The two younger groups were tested in schools in Leicestershire, 10 subjects at a time, and the 18- to 19-year-olds were tested as part of a first-year Psychology laboratory class, 20 subjects at a time. All subjects listened to the same four excerpts of music, but half of the subjects in each group were told that they were hearing "contemporary classical" excerpts and the remaining half were told that they were hearing "modern progressive rock" excerpts. Measures were taken to ensure that the subjects were not aware of this deception during the course of the experiment: in some cases two experimenters tested the two groups simultaneously, and in others, although groups were tested consecutively, the room used was sufficiently insulated to prevent either group hearing the music played to the other.

##### Musical material and questionnaires.

On the basis of a previous study, described in Part 4.3,

excerpts from four musical compositions were selected which satisfied the criterion of being stylistically ambiguous in that listeners had previously been equally as to describe them with the labels "Modern Jazz" and "Rock Jazz" as they were to describe them with the labels "Classical" and "Modern Classical".

One excerpt was taken from each of the following compositions:

"Forms and Sounds" by Ornette Coleman.

"A Leicester Jazz Suite" by David Hargreaves.

"Chamber Concerto for 13 Instruments" by Ligeti.

"Entr'acte 11: Vergine Bella" by Charles Wuorinen, from Percussion Symphony, 1976.

Each excerpt lasted 40 seconds, and the four were recorded on a cassette tape in the order shown above, with a 30-second pause between excerpts.

Subjects were required to give affective and evaluative ratings of each excerpt by completing questionnaires which consisted of four pairs of 5-point rating scales, one pair applying to each excerpt. One rating scale in each pair referred to the affective components of a listener's reaction and ranged from "dislike very much" to "like very much"; the other was evaluative, ranging from "very low quality" to "very high quality". The rating scales were presented as shown in Appendix 5.3.1 except in the case of the youngest group of subjects (10 to 11 years), where the paper was cut into four

strips, one referring to each excerpt of music, and made into a four-page booklet. It was hoped that this would help to avoid confusion, because subjects could easily see, and were told, that the first page referred to the first excerpt, the second page to the second excerpt, and so on.

#### Procedure.

Once a group of subjects had been assembled, they were given one questionnaire each. The 18- to 19-year-olds were tested in two groups of 20, and members of the other age groups were tested in four groups of 10. In each age group, half the subjects tested (i.e. two groups of 10, or one group of 20 in the case of 18- to 19-year-olds) were given the following instructions verbally:

The aim of this experiment is to find out how much people like a piece of modern progressive rock music when they first hear it; and also to find out something separate, which is how good they think its musical quality is. For instance you might like a piece of music very much but be aware at the same time that it is not of a very high quality: perhaps a bit "trashy". On the other hand, you might admire a composer's technical and creative ability and think that the quality of the music is very high, but be left totally cold by it emotionally, and not like it at all. I am particularly interested in finding out what people think about unfamiliar progressive rock music, so I have chosen four

excerpts from pieces by some modern rock groups you probably will not have heard. If you do know, or think you know, any of the pieces, please make a note of which ones on your answer-sheet. Please listen to each one very carefully. After it has finished, please answer the two questions about it on your sheet. As you can see, under the heading PIECE 1 the first question is, "How much do you like it?" and the second question is, "How good is the quality of the music?". There will be a pause long enough for you to tick whichever boxes you feel apply to you for the first excerpt, then you will be asked to do the same for the second, third and fourth excerpts. Do not try to fill in your answers while the music is playing, because there will be plenty of time between each excerpt. It is your first impressions of modern progressive rock that are important, so you will not be hearing any other type of music, and you will hear each excerpt just once.

The other 20 subjects in each age group were given identical instructions except that "modern progressive rock" was replaced by "contemporary classical" each time it appeared. As previously pointed out, steps were taken to ensure that subjects who were told they had heard "modern progressive rock" music and those who were told they had heard "contemporary classical" music did not discover that they had heard identical excerpts until after all subjects had been tested. It was stressed that "quality", in this context, was not meant to refer to the



quality of the recording or cassette player but to the quality of the music itself. Instructions were explained in more detail if necessary during the testing of 10- to 11-year-old subjects.

After subjects had rated all four excerpts they were asked to write on the back of the response sheet their age, sex and the nature and extent of any musical training they had received. Response sheets were then collected by the experimenter and marked "modern progressive rock" or "contemporary classical" according to which set of instructions had been used.

### Analysis of Responses and Results

Response sheets were initially examined in order to find out whether any of the excerpts used had been more familiar to one subgroup of subjects than another, because this might affect the interpretation of any other differences between the groups. However, it was found that no subjects in any age group had made a note of having heard any of the excerpts before, so it was assumed that they were equally unfamiliar to all subjects.

Each subject's ratings were converted to scores by ascribing numerical values from one to five for ratings of "dislike very much" to "like very much" and "very low quality" to "very high quality" respectively. Ratings for each subject were averaged over the four excerpts to give two means; one was the mean affective rating and the other was the mean evaluative rating. Subjects were subdivided according to age and also according to whether they were told the music was "modern progressive rock" or "contemporary classical". Two, 2-way

analyses of variance were performed, on the liking ratings and the quality ratings respectively. In each analysis, the two independent factors were age, and the nature of the stylistic information provided about each excerpt. Summary tables of the results of these analyses are shown in Tables 5.3.1 and 5.3.2 respectively, and the means are shown graphically in Figures 5.3.1 and 5.3.2.

None of the main effects reached statistical significance but the interaction between age and stylistic information was statistically significant for the "liking" ratings ( $F[2,114] = 5.13, p < 0.01$ ).

### Discussion

From Figure 5.3.1 it can be seen that liking for the excerpts appeared to decrease with age. This was not the case for evaluative ratings (Figure 5.3.2): 10- to 11-year-olds apparently rated the excerpts as lower in quality than did the other two age groups. If all age groups are considered together, excerpts appeared to be better liked when described as "modern progressive rock" than when they were described as "contemporary classical" music (mean liking ratings were 2.87 and 2.68 respectively). The reverse tendency appeared so far as evaluative ratings were concerned: subjects to whom excerpts were described as "modern progressive rock" gave lower ratings of their quality than those who were told the excerpts were "contemporary classical" music (the means were 2.65 and 2.75 respectively). However, none of these effects was statistically

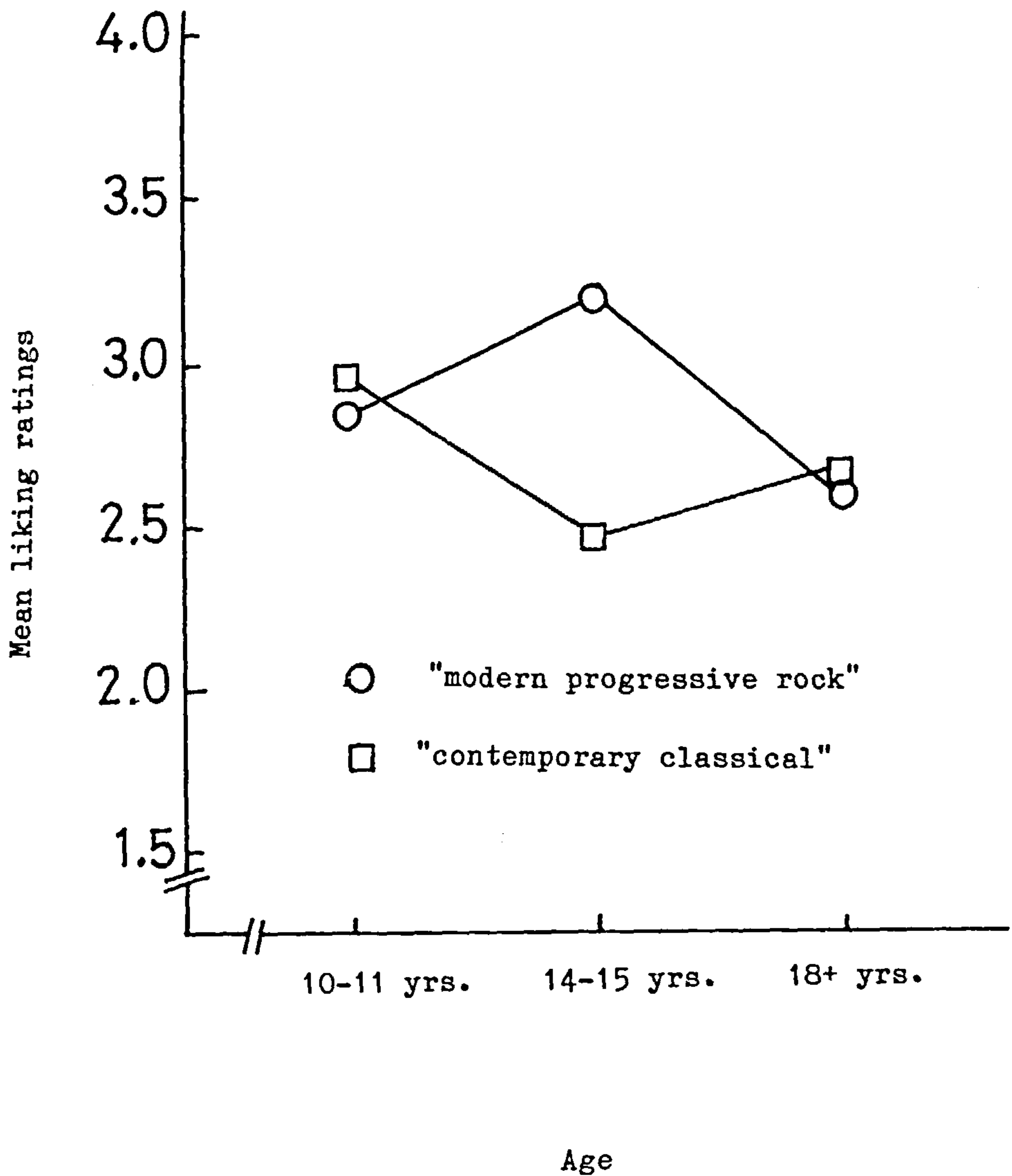


Figure 5.3.1: Liking ratings for excerpts of music described as either "contemporary classical" or "modern progressive rock", over three age groups. (1 = dislike very much, 5 = like very much).

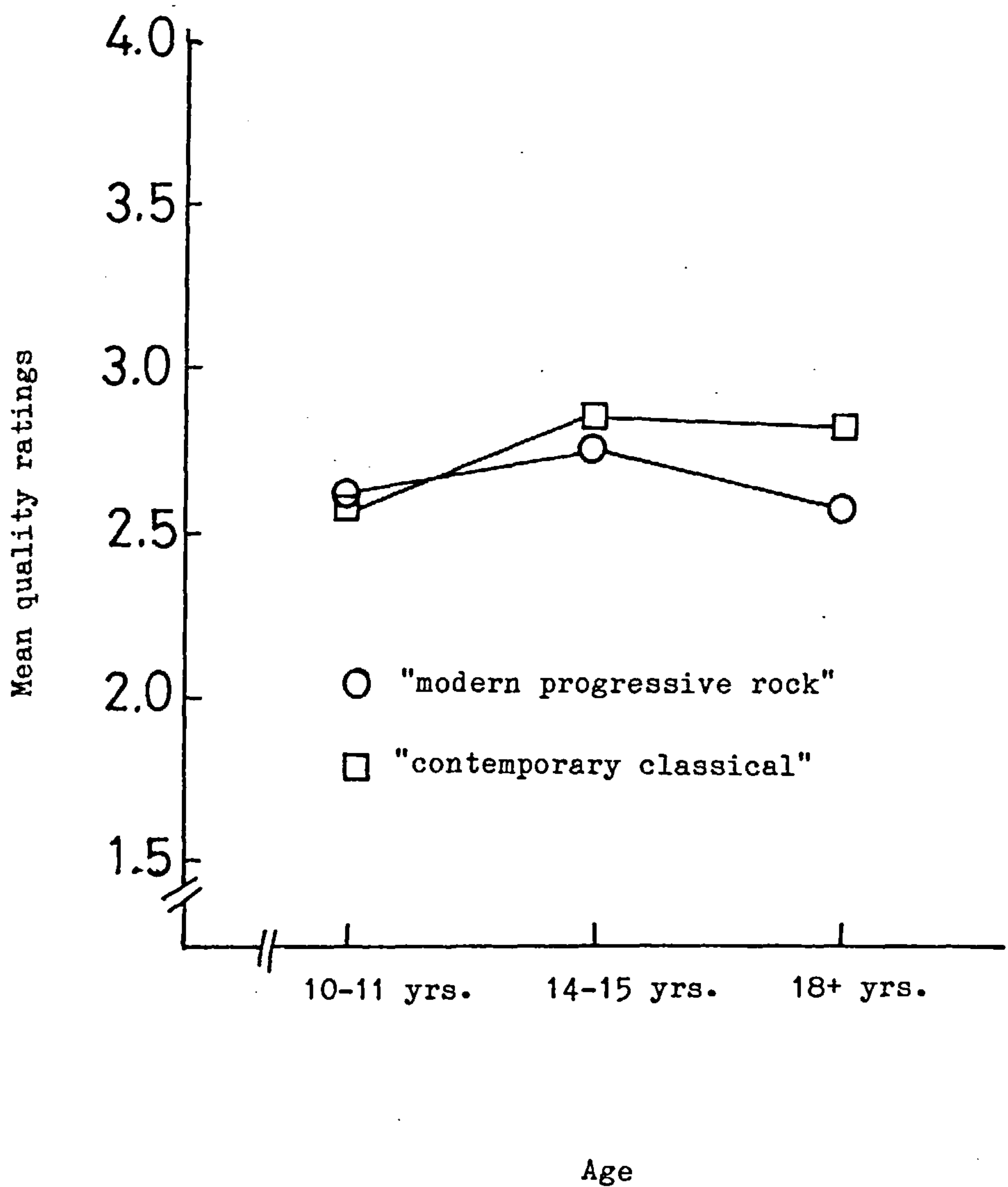


Figure 5.3.2: Quality ratings for excerpts of music described as either "contemporary classical" or "modern progressive rock", over three age groups. (1 = very low quality, 5 = very high quality).



Table 5.3.1

Analysis of variance summary table, N=120  
The effects of age and information about the source of excerpts of music on liking ratings.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	2	0.819	1.82	n.s.
B	1	1.102	2.45	n.s.
AB	2	2.344	5.22	<0.01
ERROR	114	0.449		

A=age, 10-11 years, 14-15 years, and 18-19 years  
(independent factor).  
B=type of information given as to source of excerpts,  
"modern progressive rock" or "contemporary classical"  
(independent factor).

Table 5.3.2

Analysis of variance summary table, N=120  
The effects of age and information about the source of excerpts of music on quality ratings.

Source of Variance	d.f.	Mean Square	<u>F</u> -ratio	<u>p</u>
A	2	0.851	1.27	n.s.
B	1	0.326	0.97	n.s.
AB	2	0.191	0.57	n.s.
ERROR	114	0.334		

A=age, 10-11 years, 14-15 years, and 18-19 years  
(independent factor).  
B=type of information given as to source of excerpts,  
"modern progressive rock" or "contemporary classical"  
(independent factor).

significant, so it would be a mistake to draw definitive conclusions from them. One interesting point is that the 10- to 11-year-olds seemed to display a greater fragmentation of their ratings than did the older groups: by comparing Figures 5.3.1 and 5.3.2 it can be seen that there is a larger difference between affective and evaluative ratings in this age group than there is in the other age groups.

Bearing in mind the fact that the interaction between age and stylistic information, for liking ratings, was the only statistically significant effect found, the results shown in Figures 5.3.1 and 5.3.2 may be summarised as follows.

So far as affective ratings were concerned, neither the 10- to 11-year-olds nor the 18- to 19-year-olds appeared to be greatly influenced by information as to the style of the music: within each age group, ratings were very similar regardless of whether subjects were told the music was "modern progressive rock" or "contemporary classical". There was however, a nonsignificant tendency for the 10- to 11-year-olds to give higher liking ratings overall than the 18- to 19-year-olds. The 14- to 15-year-olds, by contrast, were influenced to a statistically significant extent by the stylistic context in which the excerpts were presented: the group who were told that the excerpts were "contemporary classical" liked them considerably less than the group who were told that they were all "progressive rock" (mean affective ratings were 2.45 and 3.20 respectively).

So far as evaluative ratings were concerned, the 10- to 11-year-olds were again apparently unaffected by the context,

giving almost identical ratings in the two conditions. The 14- to 15-year-olds gave slightly higher quality ratings overall than the younger group and tended to judge the music as being of higher quality if it was described as "contemporary classical" than if it was described as "modern progressive rock". This tendency was displayed to a greater extent in the case of the adults, where separation between the two conditions became more apparent. The group who were told the excerpts were "contemporary classical" gave higher evaluative ratings than the group who were told they were "modern progressive rock".

The directions of these trends concur with the findings of Chapman and Williams (1976) and could speculatively be subsumed within the same theoretical framework. Despite the fact that in the present experiment the prestige of the excerpts was not the independent variable, no prior measure of musical preference having been taken, the finding that context influenced affective ratings in only the adolescent group (14- to 15-year-olds) is suggestive, because according to Chapman and Williams (1976) it is likely to be in this age group that the importance of conforming to peer group norms as regards musical taste is paramount. Previous research (e.g. Murdock and Phelps, 1972), as well as everyday experience, suggest that adolescents are likely to prefer popular to "serious" music.

Stylistic information did not affect subjects' evaluative ratings. However, there did appear to be a nonsignificant trend in the 18- to 19-year-old group, and the direction of this trend was that predicted on the basis of Hargreaves, Messerschmidt and Rubert's (1980) research, in that the group of subjects who were



told the excerpts were "contemporary classical" gave them higher quality ratings than those who were told they were "modern progressive rock". This is interesting because the excerpts judged by the two groups were of course identical. This finding could be interpreted by suggesting that adults believe that rock music is of a lower quality than classical music; but the issue is a complicated one because it is quite likely that "quality" may have a different meaning when applied to rock, as opposed to classical, music. Hargreaves et al. found greater separation between evaluative ratings for actual classical and popular excerpts in the case of musically trained subjects than in those with no training, and they suggested that the greater musical experience of the trained group gave them greater discernment in their quality ratings. It could be argued that in the present experiment the 18- to 19-year-olds might by virtue of their age alone be more musically experienced, and hence more discerning, than younger subjects, and that this might constitute one explanation of why it was only in this age group that there seemed to be a difference in quality ratings between the "contemporary classical" and "modern progressive rock" conditions. However, when it is remembered that both groups of subjects heard the same pieces the argument is, if anything, reversed: the older subjects appeared to be less discerning in that they were the only ones who allowed their evaluative ratings to be influenced by outside information. Rather, indirect support is given here to Hargreaves et al.'s alternative explanation of the phenomenon; they suggest that it is a product of cultural stereotyping because,

...musical training is usually classical training; these subjects have presumably had more exposure to classical music, and to a musical educational establishment that typically dismisses popular music as trivial and worthless. They equate the language of classical music with high quality....(p. 15).

It should be pointed out that the 18- to 19-year-old subjects in one respect constituted a less representative sample than either of the other age groups because they were all university undergraduates, whereas only a proportion of the 10- to 11- or 14- to 15-year-old subjects would be likely to continue their education beyond the age of 16.

Whatever the explanations of the findings with respect to affective and evaluative ratings, it seems clear that aesthetic responses to music are not made in what Konecni (1979) described as "a social-emotional vacuum". This was also apparent in the results of Experiment 5, where the stimuli used were not pieces of naturalistic music such as those used in the present experiment. It is very striking that factors which are influential in determining musical preference and quality ratings may be culture-based to an overwhelming extent and could therefore potentially be affected by educational practice: for instance, it is quite plausible that it is precisely an emphasis on classical music in education that contributes to the importance of rock music in adolescent subculture (traditionally thought of as being a rebellious subculture), and as already suggested by Hargreaves et al. (1980), popular music may be

dismissed by the musical educational establishment as trivial. In some respects therefore, it appears that stylistic labels such as "contemporary classical" and "modern progressive rock" are value-laden, as well as being purely descriptive, at least from around adolescence. The 10- to 11-year-olds appeared to be relatively unaffected by the stylistic labels, so far as their preferences and quality ratings were concerned. Because any form of specialist music education does not usually begin until after this age, it is tempting to use this as secondary evidence for the possibility that the musical educational establishment itself may be one of the causative agents in the production of a value-laden interpretation of stylistic labels. Certainly Burnett (1977), writing about secondary schools, makes the point that:

The Newson Report, DES Pamphlets, Schools Council Working Papers, all testify to the fact that music is one of the most unpopular subjects and that provision for it is poor, and they contrast this with the vitality of teenage musical life outside school.(p. )

This argument is perhaps a little far-fetched, however, in the context of the present experiment. To summarise, the important conclusions to be drawn are that adolescent subjects, like those in Chapman and Williams' (1976) study, were markedly influenced by information about the style of excerpts of music when asked to say how much they liked them: those who were told the music was "progressive rock" liked it better than those who

were told it was "contemporary classical". Ratings of the quality of the music, on the other hand, remained unaffected. There was a nonsignificant trend for the adult subjects (18- to 19-year-olds) to give slightly higher quality ratings if they were told the excerpts were "contemporary classical" than they did if they were told the excerpts were "modern progressive rock", but their affective ratings were not influenced by stylistic information. The pre-adolescent subjects (10- to 11-year-olds) were not influenced by stylistic information whether they were giving affective or evaluative ratings.



## PART 6 : GENERAL CONCLUSIONS

### 6.1: Summary of Findings

The results of the first two experiments reported in this thesis suggested that musical style sensitivity, as operationalised here, increases with age. Both experiments employed naturalistic music rather than experimental stimulus sequences, and it was found that the style of the music (i.e. whether it was popular or classical) influenced subjects' responses in several important ways. First, subjects tended to give more correct responses (i.e. to be more sensitive to stylistic differences) when judging popular, than when judging classical music. Second, there was a slight tendency for subjects to give more written justifications for their responses to classical than to popular excerpts. Third, subjects gave justifications which emphasised different stimulus characteristics according to the style of music: tempo and rhythm were stressed when judging popular excerpts, and instrumentation was stressed when judging classical excerpts. It therefore appears that the style of music used as an aesthetic stimulus in research may have a considerable effect on the results of experiments. Musical style is a variable which has been somewhat neglected by previous researchers, and the present results suggest that it is worthy of considerably more investigation.

In Experiments 3 and 4 a further exploration was undertaken of the mechanisms involved in the perception of differences between excerpts of music. The emphasis was upon the effects of stimulus characteristics on similarity judgements, and the potentially salient characteristics which were manipulated were selected on the basis of subjects' comments in Experiments 1 and 2. Experiment 3 included a comparison among age groups which revealed that subjects aged 7 to 8 years tended to notice changes in rhythmic complexity slightly less often than did older subjects (aged 10 to 11, 13 to 14, and 18 years). The 10- to 11-year-olds appeared to be particularly sensitive to alterations in tempo, but alterations in melodic complexity did not affect the four age groups' responses differentially. Overall, however, there were no clear-cut developmental trends, and it was argued that this may have been a reflection of the nonverbal response mode adopted. A possible implication of this finding is that the overall improvement in style sensitivity with age which was observed in Experiments 1 and 2 did not result merely from an improvement in the ability to detect similarities and differences among individual stimulus characteristics.

One of the most important findings of Experiments 3 and 4 was that subjective and objective measures of musical stimulus characteristics (such as rhythmic complexity and melodic complexity) did not necessarily correspond, either in magnitude or direction. Although some previous workers (Walker, 1973; Heyduk, 1975) have commented upon the importance of the differences between subjective and objective measures of

complexity, there are others (McMullen, 1974; McMullen and Arnold, 1976; Konecni and Sargent-Pollock, 1976) in which the authors base their conclusions on manipulations of objective complexity alone. There is nothing intrinsically wrong in adopting this latter approach, but a more complete understanding of the notion of musical complexity will result from acknowledging that it cannot be defined in terms of stimulus characteristics alone. The results of Experiment 5 further corroborated the notion that stimulus characteristics alone are not sufficient predictors of subjects' responses to music, because it was found that the pattern of affective responses to the same stimuli could be altered merely by asking subjects to make choices while pretending to be angry.

Musical style emerged as a variable which influenced listeners' responses in Experiments 1 and 2, and the results of Experiments 3, 4 and 5 suggested that it is possible that perception of musical style amounts to more than mere perception of similarities and differences which are described in terms of stimulus characteristics alone. However, these experiments did not directly investigate any aspect of the listeners' conceptions of musical style, so Experiment 6 was performed in order to find out how subjects' classifications of music as belonging to one style or another developed with age. Despite some flaws in the design of this experiment, the first of which was the absence of any control for order effects, the second (which emerged in retrospect) being that the stylistic labels provided were out of date, there did appear to be a developmental trend in the way stylistic labels were used to

describe excerpts of naturalistic music. This did not consist of a simple increase with age in the number of labels subjects were ready to use, nor of a decrease with age in the likelihood that inconsistent labels (such as "Traditional Folk" and "Modern Jazz") would be used together. Instead, it was apparent that 11- and 12-year-old subjects were willing to use many stylistic labels but that they did not apply them consistently; that 14- and 15-year-olds, perhaps in their eagerness to avoid inconsistency, used fewer labels than the youngest subjects; and that 18- and 19-year-olds used more labels than the youngest subjects, but in a more systematic and consistent way. It might therefore be concluded that for 11- and 12-year-olds, the concept of a musical style is rather poorly defined and has not received much consideration, whereas 14- and 15-year-olds are more aware of the existence of stylistic distinctions and are perhaps anxious to make decisions about stylistic category membership on a "one piece - one style" basis. Subjects over the age of 18 appear more willing to admit that the matter is not necessarily clear-cut, and that the boundaries between one musical style and another are not rigid.

Developmental aspects of the perception of musical style were investigated further in Experiment 7. Here it was found that the affective and evaluative reactions of 10- and 11-year-olds remained relatively unaffected by being told that some excerpts of music were either of one style or another, whereas 14- and 15-year-olds who had been told that the music was "modern progressive rock" liked it better than other subjects of the same age who had been told that the same compositions were



"contemporary classical" music. Adult subjects, in common with the 10- to 11-year-olds, did not exhibit different patterns of preference according to the information they had received about the style of the compositions. Subjects in this experiment were also asked to give ratings of the quality of the compositions. These ratings were not affected by stylistic information, although there was a nonsignificant tendency for the adults who had been told that they were hearing "contemporary classical" music to give higher quality ratings than did the adults who had been told that the music was "modern progressive rock".

In Part 1.2 it was stressed that, in view of the complexity of the area, music psychology should be studied using a variety of approaches. This has been attempted in the present thesis, in which the underlying theme has been an examination of some aspects of musical style, either directly as in Experiments 1, 2, 6 and 7 or indirectly via an examination of specific stimulus characteristics as in Experiments 3, 4, and 5. Results suggest that the perception of stylistic distinctions in music is not solely a function of the perception of individual physical characteristics of the music. Despite the finding that even adults are by no means unanimous in their overt description of music as belonging to one or another stylistic category, differential responses to different styles of music were found in subjects as young as 7 years old. This and other evidence suggest that musical style merits more consideration by researchers than it has so far received, and the last section of this thesis makes some suggestions as to ways in which the concept might be approached.

## 6.2: Implications and Suggestions for Further Research:

### The Concept of Musical Style

The findings of this thesis have potential relevance for, and may be discussed in terms of, other research in developmental, cognitive, and social psychology. Starting with a developmental approach, it was pointed out in Part 2.1 that many researchers in music psychology (Gardner, 1973; Jones, 1974; Larsen, 1973; Pflederer, 1964 and 1967) have tried to assimilate musical development into a Piagetian framework. They have argued that developing children might go through qualitatively different stages in the way in which they think about and respond to music; and that the differences between these stages might correspond to the differences between the pre-operational, concrete operational, and formal operational levels described in Piagetian research. It was pointed out in Part 2.1 that both Gardner (1979) and Serafine (1980) argue that such an approach may not be entirely appropriate so far as musical development is concerned: Serafine suggests that data obtained to date are such that it is not necessary to resort to an explanation in terms of Piagetian constructs, and Gardner argues that Piaget's view of the end-point of development does not include domains other than that of logical-rational thought. In other words, Piaget's theories refer almost exclusively to the development of scientific and mathematical reasoning, and are therefore not wholly relevant to the development of aesthetic sensitivity. Gardner suggests that, whereas Piaget

did not discuss in detail the means by which problems were presented to children, future researchers would be wise to attend to the characteristics of the particular symbol systems (whether they involve language, visual art, mathematical concepts, or music, for instance) involved in each presentation. This argument seems especially relevant to the study of responses to music, because such responses start occurring very early in life, even before a child has acquired any verbal ability, and continue to have emotional and physical expression after a person has learned to use language (Moog, 1976). Therefore, as already stressed in Part 2.1, if purely linguistic media are involved in the forms of presentation and measures of response employed in musical tasks, researchers' understanding of developmental aspects of the cognitive and emotional factors playing a part in music perception may well be incomplete and/or distorted.

The results of experiments reported in this thesis support this suggestion because in the tasks requiring a relatively high level of linguistic sophistication (e.g. Experiments 1 and 2, question [d]; the supplementary question in Experiment 3 [Table 3.3.8]; and Experiment 6), there were more, and apparently stronger, age related effects than there were in the tasks where responses were less dependent upon linguistic ability (e.g. the similarity rating task in Experiment 3).

The initial aim of the experiments described was to assess the extent to which subject's responses reflected similarities and differences within and between pieces of music. However, it gradually emerged that a unifying factor in the description of

their responses, partly dependent upon linguistic ability, was the concept of musical style. The definition of musical style poses problems: the concept has not often been investigated by music psychologists, and when it has been investigated it is quite often referred to in the same way as that in which parameters such as tempo and melody are mentioned: that is, it is referred to as if it were a relatively objective feature of a composition which is dependent upon physical characteristics of that composition. The kind of definition of musical style which might arise from previous studies (Greer et al., 1974; LeBlanc, 1981) is one based on the rhythmic, melodic, and tonal qualities of a set of compositions which are such that they distinguish that set of compositions from another set. LeBlanc (1981) specifically states that,

Style is considered a physical property of music because a composer's adherence to a particular one restricts the music devices available at a given point. This restriction is especially evident in the popular styles, with their traditions of tempo and performing medium. (p. 143).

Gardner (1973) chose to give an operational definition of style sensitivity rather than attempting to describe style in an abstract sense, and in so doing he avoided the problem of having to define how children are able to distinguish between one style and another. There are many good reasons for considering style as a physical property of music, and most people would probably try to define style in physical terms, if asked. However, the



findings of this thesis and of Chapman and Williams (1976) suggest that decisions as to the style of a piece of music may not be dependent on physical stimulus properties alone. In other words, there may be an important subjective element in the attribution of style. The results of Experiments 6 and 7 show that the subjective element of style labelling may be particularly noticeable in those cases where the physical properties of the music are such that excerpts cannot be easily classified: this might be predicted by common sense. However, it does not logically imply that the subjective element is lacking for those compositions which can be more easily labelled. In order to pursue this line of argument further, it should be established whether or not people are able to detect similarities and differences between ambiguous excerpts such as those used in Experiments 6 and 7 with as much ease as they can detect differences between less ambiguous excerpts like those used in Experiments 1 and 2, using an experimental paradigm similar to that used in Experiments 1 and 2. It would also be interesting to find out, using the technique described in Experiment 6, whether subjects would be unanimous in their descriptions of the styles of excerpts which were not selected on the basis of their ambiguity.

There are many ways in which the developmental aspects of the concept of musical style could be investigated further. In Part 4.1 it was suggested that decisions about musical style might be analogous in some respects to decisions about category membership. If this were the case, findings about the development of categorisation in children (e.g. Rosch, 1978)

would provide a new source of information relevant to the development of music perception. One suggestion made in Part 4.1 was that it would be interesting to find out whether descriptions of music would become more specific with increasing age.

The linguistic aspects of the perception of style were stressed in Experiment 6 in a task which all subjects, even the adults, found extremely difficult. This might imply that it is not a usual response for people to apply a verbal stylistic label to the music to which they listen; however, the range of labels provided in this experiment may have been excessively restrictive. It would therefore be necessary to find out whether subjects would experience the same difficulty in labelling the excerpts even if they were allowed to use any label they wanted, before concluding that it is not an immediate reaction so far as most people are concerned to decide what stylistic labels are applicable to the music they hear. In addition, the same open-ended task should be carried out using a wider range of musical stimuli than that presented in Experiment 6 if any general conclusions are to be drawn about the development of the ability to label musical styles.

Overall, with respect to the developmental issues raised in this thesis, it may be said that Machotka's (1966) "Piagetian" predictions about the genesis of the concept of style in the visual arts are not entirely appropriate to music, especially because his argument that a child can conceive of several styles of representation only after s/he has reached the hypothetico-deductive level of thought depends on the representational

nature of visual art; and music is not representational in the same sense as drawings or paintings, if at all. Children as young as 7 or 8 years old are apparently quite capable of distinguishing between musical styles, although they have difficulty in expressing these distinctions verbally (Experiment 2), and the way in which stylistic sensitivity was found to change with age (Experiments 1 and 2) did not seem to call for an explanation in terms of Piagetian theory. Gardner's (1979) argument that any model of aesthetic development must take into account the characteristics of the symbol system involved in the art in question seems clearly appropriate here.

From a cognitive point of view, one thing which becomes apparent from the research described in this thesis is that it is important, when discussing style, to make clear whether one is referring only to physical similarities and differences within music without labelling it as one style or another (as in Experiments 1, 2, and 3), or whether style is being discussed with reference to particular verbal descriptions of different styles. In the latter case the results of Experiment 7 suggest that it is a distinct possibility that factors other than stimulus characteristics will automatically become salient to the subjects involved. In addition, examination of the age related effects in Experiments 1 and 2, contrasted with their absence in Experiment 3, suggest that the perception of stylistic similarities and differences (without explicit verbal labels) amounts to more than the perception of the sum of individual variations in physical stimulus characteristics. All these findings taken together lead to the suggestion that one

way to understand the concept of musical style might be in terms of a hierarchy with three levels. At the first level are individual stimulus characteristics, which give rise to (but are not exactly equivalent to) implicit, more subjective, stylistic characteristics such as those investigated in Gardner's (1973) research and in Experiments 1 and 2. These implicit stylistic characteristics represent the second level of the hierarchy. These in turn give rise to (but are again not exactly equivalent to) explicit stylistic labels such as those whose use was examined in Experiment 6. It could be argued that the difference between the first and the second level can be described in terms of an accumulation of characteristics which have a subjective component as well being definable in purely objective terms. An example of such a characteristic might be complexity: it was shown in Experiments 3, 4, and 5 that although complexity can be measured by objective means, perceived complexity is not determined solely by physical stimulus characteristics and may alter independently of these (e.g. in accordance with an imagined emotional state). It could be argued further that the way in which the second level of the hierarchy gives rise to the third might be via language and categorisation; thus subjective features (perceiver characteristics) come into play to an even greater extent here because the way in which music is classified verbally may vary considerably from person to person (Experiment 6).

There may be age related effects at all levels of this hierarchy: in the present thesis, Experiment 3 investigated developmental trends at the first level, Experiments 1 and 2



investigated developmental trends at the second level, and Experiments 6 and 7 were concerned with the third level.

It is at the third level of the hierarchy that factors other than cognitive or developmental ones become particularly relevant, because it is at this level that the concept of style becomes most subjective and therefore most amenable to extramusical influences. For example, in Experiment 7 it was found that a change in a verbal label alone was sufficient to alter some subjects' affective ratings of excerpts of music. There was an interaction between subjects' ages and the extent of the effect of an alteration in the description of the music: adolescents were the only subgroup whose ratings were affected. It would be interesting to find out whether this age group was unique in this respect by testing children younger than those in the present sample.

One explanation of the findings of Experiment 7 was in terms of the importance of peer group norms to adolescents. This introduces the question of the extent to which social influences play a part in responses to music, and leads to a suggestion for further research. It would be illuminating to find out whether the adolescents in Experiment 7 would have responded differently had they been asked to make their ratings individually: although they did not confer during testing, it is possible that each subject may have been aware that his or her neighbour could see his or her response sheet with little effort. Preliminary results from research such as that by Chapman and Williams (1976) and Konecni (1979) suggest that social factors are worthy of consideration, and the results of

Experiment 7 further suggest that the importance of such factors as determinants of responses to music might alter with age.

In summary, it has been proposed that responses to music and to musical style in particular might be described in terms of a three-tiered hierarchical conception of musical style. At the first level the emphasis is upon stimulus characteristics such as pitch and tempo which may be completely defined in terms of their physical properties. These characteristics in combination produce phenomena such as complexity which have a definite subjective component, even although they may also be defined in physical, objective terms. Implicit stylistic differences emerge between compositions via such phenomena; and these stylistic differences, although they depend ultimately upon stimulus characteristics, are not as easy to describe objectively as are differences at the first level of the hierarchy. At the third level of the hierarchy, the emphasis is upon verbal descriptions of the implicit stylistic differences present at the second level, and this level is therefore the most subjective of the three.

It is suggested that this way of understanding musical style is useful in that it provides a means of accounting for the finding that responses to changes in individual stimulus characteristics (as in Experiment 3, which embodied a first level task) did not alter with age in the same way as did responses to natural music in Experiments 1 and 2 (which embodied a second level task), and that "style" emerges as important when investigated at the second level (Gardner, 1973; Greer et al., 1974; Experiments 1 and 2 in this thesis) whereas

it does not necessarily seem to be important to subjects to find labels for musical styles, which constitutes a third level task (Experiment 6). Not only does the three-tiered model help to account for the present findings, but it also offers a tentative conceptual framework upon which further research might be based: a different set of implicit assumptions is associated with each level, and therefore hypotheses about responses to music can be generated which are specific to each level. For example, it might be fruitful to examine further the differences between developmental trends in responses to level one, level two, and level three tasks. The issues surrounding the concept of musical style are complicated, and the area has been explored very little, but it is to be hoped that the proposed model will provide a useful starting point for further research.

## APPENDICES



Explanation of terms and symbols used in Appendices 3.3.2,  
3.3.3, 3.3.4, and 3.5.1.

Dotted vertical lines separate one beat from the next.

Solid vertical lines separate one bar from the next.

The terms "high hat", "accent", etc. refer to possible

options on the rhythm generator. Each occasion upon

which each option is used is marked with an "x", and

all "x's" in the same vertical column occur

simultaneously. Accents are not counted as "events"

because they do not contribute a separate sound; they

emphasise whatever sound is being produced in conjunction

with them.

### Appendix 2.3.1

The "practice pair" used in Experiments 1 and 2 consisted of one excerpt taken from each of the following compositions:

"Lord of the Rings" by Bo Hansson.

"Walking on the Moon" performed by the Police.

If the classical compositions used as sources of the stimulus material are represented by the letters A, B, and C, standing for the compositions by Ravel, Mozart, and Munrow respectively; and if the popular compositions used are represented by the letters a, b, and c, standing for the compositions by The Modern Jazz Quartet, Pink Floyd, and Focus respectively, then the orders in which pairs of excerpts appeared on each of the two stimulus tapes are as follows:

TAPE 1: A-B, A-C, a-a, B-B, A-A, b-b, b-c, a-c, c-c, a-b, C-C,  
B-C.

TAPE 2: B-B, a-a, b-b, a-b, A-B, C-C, b-c, A-C, c-c, B-C, A-A,  
a-c.

Appendix 2.3.2

Page from questionnaire (Experiments 1 and 2).

Do you think that they are from the same piece?

\_\_\_\_\_

If not, could the same person have composed them?

\_\_\_\_\_

Have you heard either before?

NO      1st      2nd

What made you decide whether or not they were from the same piece?

---

Do you think that they are from the same piece?

\_\_\_\_\_

If not, could the same person have composed them?

\_\_\_\_\_

Have you heard either before?

NO      1st      2nd

What made you decide whether or not they were from the same piece?

---

Do you think that they are from the same piece?

\_\_\_\_\_

If not, could the same person have composed them?

\_\_\_\_\_

Have you heard either before?

NO      1st      2nd

What made you decide whether or not they were from the same piece?

---

Do you think that they are from the same piece?

\_\_\_\_\_

If not, could the same person have composed them?

\_\_\_\_\_

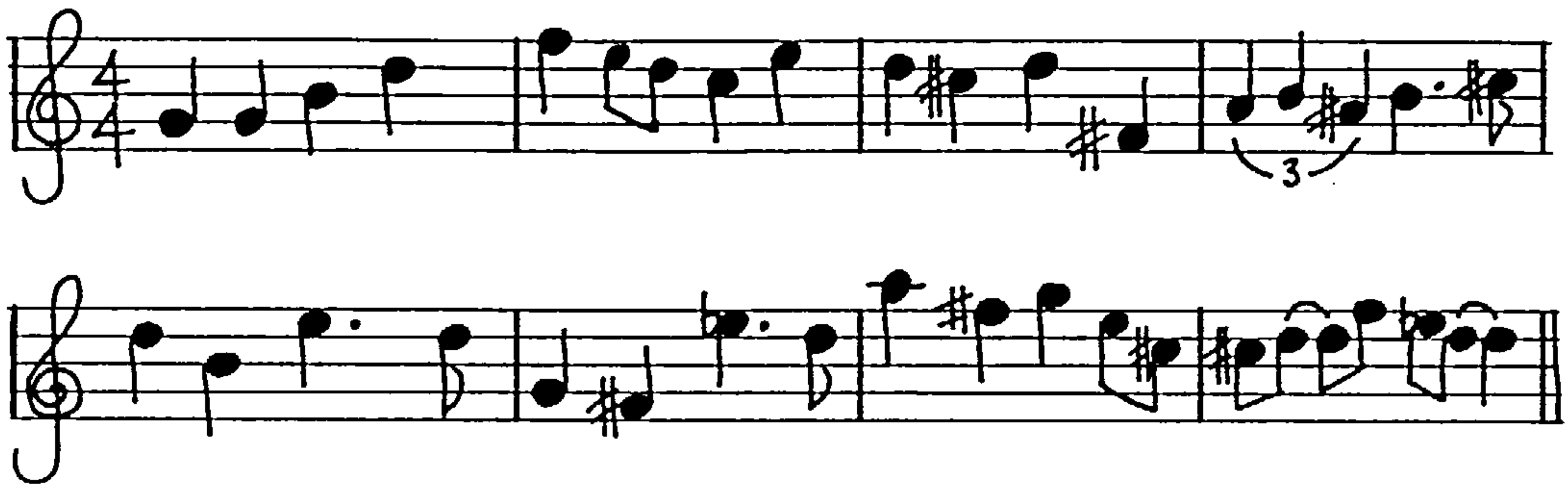
Have you heard either before?

NO      1st      2nd

What made you decide whether or not they were from the same piece?

### Appendix 3.3.1

The melodic component of the standard sequence used in Experiments 3 and 4.





### Appendix 3.3.2

The rhythmic component of the standard sequence used in Experiments 3 and 4.

Rhythm for bars 1, 3, 5, and 7.

High hat	X	X	X	X	
Accent	X		X		
Rim shot		X		X	
Snare					
Bass drum	X	X	X	X	

Rhythm for bars 2, 4, 6, and 8.

High hat	X	X	X	X	
Accent	X		X		
Rim shot		X		X	
Snare					
Bass drum	X	X	X	X	

Mean number of events per bar = 10.5

Proportion of events which are "on the beat" = 0.76

Proportion of events which are "off the beat" = 0.24

### Appendix 3.3.3

The rhythmic component of variation three (more complex than standard) used in Experiments 3 and 4.

Rhythm for bars 1, 3, 5, and 7.

High hat	X		X		X		X		X		X		X		X	
Accent	X			X			X			X			X			X
Rim shot									X		X				X	
Snare	X			X			X			X			X			X
Bass drum	X	X		X	X						X					

Rhythm for bars 2, 4, 6, and 8.

High hat	X		X		X		X		X		X		X		X	
Accent			X			X			X			X			X	
Rim shot			X			X			X			X			X	
Snare									X		X				X	
Bass drum	X						X							X	X	X

Mean number of events per bar = 21.5

Proportion of events which are "on the beat" = 0.37

Proportion of events which are on "half-beats" = 0.42

Proportion of events which are between "half-beats" = 0.21

Appendix 3.3.4

The rhythmic component of variation four (less complex than standard) used in Experiments 3 and 4.

Rhythm for bars 1 to 8 inclusive (i.e. every bar).

High hat	X				X					X					X				
Accent	X																		
Rim shot																			
Snare																			
Bass drum	X										X								

Mean number of events per bar = 6.  
All are "on the beat".

### Appendix 3.3.5

The melodic component of variation five (more complex than standard) used in Experiments 3 and 4.





Appendix 3.3.6

The melodic component of variation six (less complex than standard) used in Experiments 3 and 4.



Appendix 3.3.7

Page from questionnaire used in Experiment 3.

How many marks does this "spare" get for being a good match?

$\frac{1}{10}$     $\frac{2}{10}$     $\frac{3}{10}$     $\frac{4}{10}$     $\frac{5}{10}$     $\frac{6}{10}$     $\frac{7}{10}$     $\frac{8}{10}$     $\frac{9}{10}$     $\frac{10}{10}$

To help the producer, can you suggest any ways of changing the "spare" to make it a better match?

### Appendix 3.4.1

A sample response sheet for giving liking ratings in Experiment 4. Response sheets for rating complexity were identical except for the first line, which read, "How complex is each piece, compared to the 'standard'?".

How much do you like each piece, compared to the "standard"?

	Much less	Less	Slightly less	Same	Slightly more	More	Much More
Piece 1							
Piece 2							
Piece 3							
Piece 4							
Piece 5							
Piece 6							
Piece 7							

3. Age: Sex: M / F

4. Please give brief details of any musical training or experience you have had, saying how long it was for.

Appendix 3.5.1

Rhythms A (low complexity), B (intermediate complexity), and C (high complexity) used in Experiment 5.

A.	High hat		X		X		X		X	
	Accent									
	Rim shot									
	Snare									
	Bass drum		X				X			

B.	High hat		X	X	X		X	X	X		X	X	X	
	Accent						X					X		
	Rim shot						X					X		
	Snare		X				X		X		X		X	
	Bass drum		X										X	

Rhythm for bars 1, 3, 5, 7, 9, and 11.

C.	High hat		X		X		X		X		X		X	
	Accent		X		X			X			X		X	
	Rim shot													
	Snare													
	Bass drum		X	X	X		X		X		X	X	X	

Rhythm for bars 2, 4, 6, 8, 10, and 12.

High hat										
Accent					X				X	
Rim shot					X				X	
Snare					X				X	
Bass drum		X	X		X		X		X	

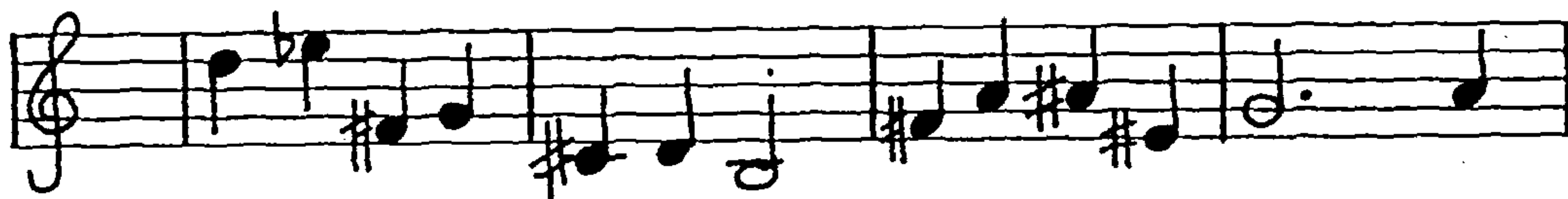


### Appendix 3.5.2

In order to obtain independent ratings of the subjective complexity of each of rhythms A, B, and C, used in Experiment 5, 54 undergraduate students were asked to listen to the rhythms and to place them in order of increasing complexity. It was found that rhythm A was judged as least complex by 51 of the students, rhythm B was judged to be of intermediate complexity by 52 of the students, and rhythm C was rated as most complex by 53 of the students. It was therefore concluded that the relative complexities of the three rhythms were perceived in the manner intended.

Appendix 3.5.3

Melodies 1, 2, and 3 which were used to accompany the rhythms in Experiment 5.



Appendix 3.5.4

Questionnaire administered to the control group in Experiment 5.

(Questionnaire 1).

- a) You will be played three extracts of music. After listening to them, please indicate which one you liked most, by ticking the relevant box below. (If you would have preferred silence, please say so).

I prefer:

1st piece	2nd piece	3rd piece	silence
-----------	-----------	-----------	---------

Can you say why? If so, give reason(s) very briefly.

- b) If you're happy and relaxed at home on your own, what type(s) of music do you most like to listen to, if any?

- c) Very roughly speaking, how do you feel now?

Very calm	Slightly irritable	Very irritable
-----------	--------------------	----------------

In order to compare your opinions with those of other people, I should be very grateful if you would mark your age-group and sex below:

Age-group:	Under 15	15-30	30-55	over 55
------------	----------	-------	-------	---------

Sex:	Male	Female
------	------	--------

Appendix 3.5.5

Questionnaire administered to the experimental group in  
Experiment 5. (Questionnaire 2).

- a) Imagine that you are feeling very angry or highly annoyed.  
Feeling like this, which of the following extracts of music  
would you most like to listen to? (If you would prefer  
silence, say so).

I would prefer: (please tick box)

1st piece	2nd piece	3rd piece	silence
-----------	-----------	-----------	---------

Can you say why? If so, give reason(s) very briefly.

- b) If you're happy and relaxed at home on your own, what  
type(s) of music do you most like to listen to, if any?

- c) Very roughly speaking, how do you feel now?

Very calm	Slightly irritable	Very irritable
-----------	--------------------	----------------

In order to compare your opinions with those of other  
people, I would be very grateful if you would mark your  
age-group and sex below:

Age-group	Under 15	15-30	30-55	over 55
Sex	Male	Female		

THANK YOU VERY MUCH

Appendix 3.5.6

Summary of the presentation of questionnaires and stimulus sequences used in Experiment 5.

In both the control group (i.e. those subjects who received Questionnaire 1) and the experimental group (i.e. those subjects who received Questionnaire 2), there were equal numbers of subjects were tested under each of the conditions shown in the cells of the table below:

	RHYTHM A	RHYTHM B	RHYTHM C
MELODY 1			
MELODY 2			
MELODY 3			



Appendix 4.3.1

Response sheet used in Experiment 6.

Please listen to these extracts of music. After hearing each one, please tick next to any name that you think might describe the type of music it is. For instance, if you feel that the first piece might be "brass band", "trad jazz", possibly "classical", then you should tick all three. It is important that you tick every category-name which you feel could describe the extract, even if you are unsure about it; there is no limit to the number of ticks you are allowed.

	1	2	3	4	5	6	7	8	9	10	11	12
Brass band												
Country and western												
Trad jazz												
Modern classical												
Punk/New Wave												
Modern jazz												
Modern folk												
Rock jazz												
Opera												
Trad folk												
Avant garde												
Muzak												
Hymn												
Classical												
Romantic												
Blues												
"Musical"												
Non-European												
Soul												

PIECE 1

How much do you like it?

Dislike very much	Dislike	Don't know	Like	Like very much
-------------------	---------	------------	------	----------------

How good is the quality of the music?

Very low quality	Low quality	Don't know	High quality	Very high quality
------------------	-------------	------------	--------------	-------------------

PIECE 2

How much do you like it?

Dislike very much	Dislike	Don't know	Like	Like very much
-------------------	---------	------------	------	----------------

How good is the quality of the music?

Very low quality	Low quality	Don't know	High quality	Very high quality
------------------	-------------	------------	--------------	-------------------

PIECE 3

How much do you like it?

Dislike very much	Dislike	Don't know	Like	Like very much
-------------------	---------	------------	------	----------------

How good is the quality of the music?

Very low quality	Low quality	Don't know	High quality	Very high quality
------------------	-------------	------------	--------------	-------------------

PIECE 4

How much do you like it ?

Dislike very much	Dislike	Don't know	Like	Like very much
-------------------	---------	------------	------	----------------

How good is the quality of the music?

Very low quality	Low quality	Don't know	High quality	Very high quality
------------------	-------------	------------	--------------	-------------------

## REFERENCES

- Andrews, M.L., & Madeira, S.S. (1977). The assessment of pitch discrimination ability in young children. Journal of Speech and Hearing Disorders, 42, 279-286.
- Asch, S.E. (1948). The doctrine of suggestion, prestige and imitation in social psychology. Psychological Review, 55, 250-276.
- Bamberger, J. (1975). The development of musical intelligence 1: strategies for representing simple rhythms. Massachusetts Institute of Technology, Artificial Intelligence Laboratory. A.I. Memo No. 342, Logo memo 19, November.
- Bartlett, D.L. (1973). Effect of repeated listenings on structural discrimination and affective response. Journal of Research in Music Education, 21, 302-317.
- Berlyne, D.E. (1967). Arousal and reinforcement. In D. Levine (Ed.), Nebraska symposium on motivation, vol. 15. Lincoln: University of Nebraska Press.
- Berlyne, D.E. (1971). Aesthetics and psychobiology. New York: Appleton-Century Crofts.
- Berlyne, D.E. (1974). Studies in the new experimental aesthetics. London: Wiley.
- Borges, J.L. (1966). Other inquiries 1937-1952. New York: Washington Square Press.
- Botvin, G.J. (1974). Acquiring conservation of melody and cross-modal transfer through successive approximation. Journal of Research in Music Education, 22, 226-233.

- Bradley, I.L. (1971). Repetition as a factor in the development of musical preferences. Journal of Research in Music Education, 19, 295-298.
- Bullock, W.J. (1973). A review of measures of musico-aesthetic attitude. Journal of Research in Music Education, 21, 331-344.
- Bullogh, E. (1921). Recent work in experimental aesthetics. British Journal of Psychology, 12, 76-99.
- Burnett, M. (ed.). (1977). Music Education Review, In defence of pop, Volume 1. London: Chappell.
- Cantor, J.R., & Zillman, D. (1973). The effect of affective state and emotional arousal on music appreciation. Journal of General Psychology, 89, 97-108.
- Cattell, R.B., & Saunders, D.R. (1954). Musical preferences and personality diagnosis. Journal of Social Psychology, 39, 3-24.
- Chapman, A.J., & Williams, A.R. (1976). Prestige effects and aesthetic experiences: adolescents' reactions to music. British Journal of Social and Clinical Psychology, 15, 61-72.
- Cleall, C. (1981). Notes toward the clarification of creativity in music education. Psychology of Music, 9 (1), 44-47.
- Colwell, R. (1969-70). The development of the music achievement test series. Council for Research in Music Education, Bulletin no. 22, 57-73.
- Conley, J.K. (1981). Physical correlates of the judged complexity of music by subjects differing in musical background. British Journal of Psychology, 72, 451-464.
- Crowther, R., & Durkin, K. (1982). Towards an applied psycholinguistic study of musical concept development. Psychology of Music, Special Issue: Proceedings of the Ninth



- International Seminar on Research in Music Education, 26-30.
- Crozier, J.B. (1974). Verbal and exploratory responses to sound sequences varying in uncertainty level. In D.E. Berlyne (Ed.), Studies in the new experimental aesthetics, London: Wiley.
- Davies, J.B. (1969). An analysis of factors involved in musical ability, and the derivation of tests of musical aptitude. Unpublished Ph.D. thesis, University of Durham library.
- Davies, J.B. (1971). New tests of musical aptitude. British Journal of Psychology, 62, 557-567.
- Davies, J.B. (1978). The psychology of music. London: Hutchinson.
- Davies, J.B. (1982). (Review of The psychology of music, D. Deutsch [ed.], Academic Press), Psychology of Music, 10 (2), 39-42.
- Deutsch, D. (Ed.). (1982). The psychology of music. Academic Press (series in cognition and perception).
- Duerksen, G.L. (1972). Some effects of expectation on evaluation of recorded musical performance. Journal of Research in Music Education, 20, 268-272.
- Eagle, C.T.Jr. (1971). Effects of existing mood and order of presentation of vocal and instrumental music on rated mood responses to that music. Doctoral dissertation, The University of Kansas, Ann Arbor, Michigan: University Microfilms no.71-27 139.
- Fay, P.J., & Middleton, W.C. (1941). Relationship between musical talent and preferences for different types of music. Journal of Educational Psychology, 32, 573-583.
- Fechner, G.T. (1876). Vorschule der Asthetik. Leipzig: Breitkopf & Hartel.



- Frances, R. (1963). Limites et nature des effets de prestige. Journ. Psychol. norm. path. 47-51, 439-457.
- Gabrielsson, A. (1982). Performance and training of musical rhythm. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education, 42-46.
- Gardner, H. (1970). Children's sensitivity to painting styles. Child Development, 41, 813-821.
- Gardner, H. (1971). The development of sensitivity to artistic styles. The Journal of Aesthetics and Art Criticism, 29, 515-527.
- Gardner, H. (1973). Children's sensitivity to musical styles. Merrill-Palmer Quarterly of Behaviour and Development, 19 (1), 67-77.
- Gardner, H. (1979). Developmental psychology after Piaget: an approach in terms of symbolization. Human Development, 22, 73-88.
- Gardner, H., & Gardner, J. (1970). Developmental trends in sensitivity to painting styles and subject matter. Studies in Art Education, 12, 11-16.
- Geringer, J.M. (1980). Verbal and operant music listening preferences in relation to age and musical training. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education, 47-50.
- Geringer, J.M., & McManus, D. (1979). A survey of musical taste in relationship to age and musical training. College Music Symposium, 19A, 69-76.
- Getz, R.P. (1966). The effects of repetition on listening

- response. Journal of Research in Music Education, 14, 178-192.
- Ghiselin, M.T. (1981). Categories, life and thinking. The Behavioral and Brain Sciences, 4, 269-313.
- Gordon, E. (1970). Iowa tests of music literacy. Iowa City, Bureau of Educational Research and Service.
- Greer, R.D., Dorow, L.G., & Randall, A. (1974). Music listening preferences of elementary school children. Journal of Research in Music Education, 22, 284-291.
- Halpin, D.D. (1943). Industrial music and morale. Journal of the Acoustical Society of America, 15, 116-123.
- Hargreaves, D.J. (1982). The development of aesthetic reactions to music. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education 51-54.
- Hargreaves, D.J., Castell, K.C., & Crowther, R. (1984). The effects of stimulus familiarity on conservation-type responses to tone sequences: A cross-cultural study. Manuscript submitted for publication.
- Hargreaves, D.J., & Colman, A.M. (1981). The dimensions of aesthetic reactions to music. Psychology of Music, 9 (1), 15-20.
- Hargreaves, D.J., Messerschmidt, P., & Rubert, C. (1980). Musical preference and evaluation. Psychology of Music, 8 (1), 13-18.
- Heinlein, C.P. (1928). The affective characters of the major and minor modes in music. Journal of Comparative Psychology, 8, 101.
- Hermanson, L.W. (1971). An investigation of the effects of

- timbre on simultaneous vocal pitch acuity of young children.
- Doctoral dissertation, Teachers' College, Columbia University.
- Hevner, K., & Landsbury, J. (1935). Oregon musical discrimination tests. Chicago: C.H. Stoelting.
- Heyduk, R.G. (1975). Rated preference for musical compositions as it relates to complexity and exposure frequency. Perception and Psychophysics, 17 (1), 84-91.
- Hoffer, C. (1982). Work related attitudes and problems of Indiana music teachers. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education, 59-62.
- Honess, T. (1981). Girls' and boys' perceptions of their peers: Peripheral versus central and objective versus interpretive aspects of free descriptions. British Journal of Psychology, 72, 485-497.
- Inglefield, H.G. (1972). Conformity behavior reflected in the musical preferences of adolescents. Contributions to Music Education, 1, 56-65.
- Johnstone, J., & Katz, E. (1957). Youth and popular music: a study in the sociology of taste. American Journal of Sociology, 62, 563-568.
- Jones, (1976). The development of the child's conception of meter in music. Journal of Research in Music Education, 24, 142-154.
- Jorgensen, E. (1980). On the development of a theory of music instruction. Psychology of Music, 8 (2), 25-30.
- Kemp, A. (1982). Personality traits of successful music teachers. Psychology of Music, Special Issue: Proceedings of

the Ninth International Seminar on Research in Music Education, 72-75.

Keppel, G., & Saufley, W.H.Jr. (1980). Introduction to design and analysis: A student's handbook. San Francisco: W.H. Freeman and co.

King, C.P. (1972). The conservation of melodic pitch patterns by elementary children as determined by ancient Chinese music. Doctoral dissertation, Ohio State University.

Konecni, V.J. (1979). Determinants of aesthetic preference and effects of exposure to aesthetic stimuli: social, emotional, and cognitive factors. Progress in Experimental Personality Research, 9, 149-197.

Konecni, V.J., & Sargent-Pollock, D. (1976). Choice between melodies differing in complexity under divided-attention conditions. Journal of Experimental Psychology: Human Perception and Performance, 2, 347-356.

Larsen, R.L. (1973). Levels of conceptual development in melodic permutation concepts based on Piaget's theory. Journal of Research in Music Education, 21, 256-263.

Leblanc, A. (1979). Generic style music preferences of fifth-grade students. Journal of Research in Music Education, 27, 255-270.

Leblanc, A. (1981). Effects of style, tempo, and performing medium on children's music preference. Journal of Research in Music Education, 29, 143-156.

Leblanc, A. (1982). An interactive theory of music preference. Journal of Music Therapy, XIX (1), 28-45.

Livesley, W.J., & Bromley, D.B. (1973). Person perception in



- childhood and adolescence. London: Wiley.
- Lowery, H. (1926). Cadence and phrase tests in music. British Journal of Psychology, 17, 111-118.
- Machotka, P. (1963). Le developpement des criteres esthetiques chez l'enfant. Enfance, 4-5, 357-379.
- Machotka, P. (1966). Aesthetic criteria in childhood: justifications of preference. Child Development, 37, 877-885.
- Madsen, C. (1982). The effect of contingent teacher approval and withholding music performance on improving attentiveness. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education, 76-81.
- McMahon, O. (1982). A comparison of language development and verbalisation in response to auditory stimuli in pre-school age children. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education, 82-85.
- McMullen, P.T. (1974). Influence of number of different pitches and melodic redundancy on preference responses. Journal of Research in Music Education, 22, 198-204.
- McMullen, P.T., & Arnold, M.J. (1976). Preference and interest as functions of distributional redundancy in rhythmic sequences. Journal of Research in Music Education, 24, 22-31.
- Meyer, L.B. (1956). Emotion and meaning in music. University of Chicago Press.
- Michael, W.B., Rosenthal, B.G., & De Camp, M.A. (1949). An experimental investigation of prestige-suggestion for two types of literary material. Journal of Psychology, 28, 303-323.
- Moffit, A.R. (1971). Consonant cue perception by twenty- to



- twenty-four-week-old infants. Child Development, 42, 717-731.
- Moog, H. (1976). The musical experience of the pre-school child.  
London: Schott.
- Murdock, G., & Phelps, G. (1972). Youth culture and the school revisited. British Journal of Sociology, 23, 478-482.
- Nelson, D.J. (1980). The conservation of metre in beginning violin students. Psychology of Music, 8 (1), 25-33.
- Parsons, M.J. (1976). A suggestion concerning the development of aesthetic experience in children. Journal of Aesthetics and Art Criticism, 34, 305-314.
- Payne, E. (1967). Musical taste and personality. British Journal of Psychology, 58, 133-138.
- Petzold, R.G. (1966). Auditory perception of musical sounds by children in the first six grades. University of Wisconsin Cooperative Research Project 1051.
- Petzold, R.G. (1969). Auditory perception by children. Journal of Research in Music Education, 17, 82-87.
- Pflederer, M. (1964). The responses of children to musical tasks embodying Piaget's principle of conservation. Journal of Research in Music Education, 12, 251-268.
- Pflederer, M. (1967). Conservation laws applied to the development of musical intelligence. Journal of Research in Music Education, 15, 215-223.
- Pflederer Zimmerman, M., & Sechrest, L. (1970). Brief focused instruction and musical concepts. Journal of Research in Music Education, 18, 25-36.

- Plummeridge, C. (1980). Creativity and music education - the need for further clarification. Psychology of Music, 8 (1), 34-40.
- Querleu, D., Renard, X., & Crepin, G. (1981). Perception auditive et reactive foetale aux stimulations sonores. J. Gyn. Obst. Biol. Repr., 10, 307-314.
- Quine, W.V. (1969). Natural kinds. In W.V. Quine, Ontological relativity and other essays. New York: Columbia University Press.
- Rasch, R.A., & Plomp, R. (1982). The perception of musical tones. In D. Deutsch (ed.), The psychology of music, (pp. 1-24). Academic Press.
- Reed, E.S. (1981). The demise of mental representations. The Behavioral and Brain Sciences, 4, 297-298.
- Report of the Fourth International Seminar on Research in Music Education (1974). Music education for the very young child. Wellington: New Zealand Council for Educational Research.
- Rigg, M.G. (1948). Favorable versus unfavorable propaganda in the enjoyment of music. Journal of Experimental Psychology, 38, 78-81.
- Rosch, E. (1978). Principles of categorization. In E. Rosch and B.B. Lloyd (eds.), Cognition and categorization, (pp.28-48). London: John Wiley & Sons.
- Rosch, E. & Lloyd, B.B. (eds.). (1978). Cognition and categorization. London: John Wiley & Sons.
- Santayana, G. (1920). Little essays, pt. (iii) no. 54. From Oxford dictionary of quotations, p.414. (1979). Oxford University Press.

- Schuckert, R.F. & McDonald, R.L. (1968). An attempt to modify the musical preferences of preschool children. Journal of Research in Music Education, 16, 39-44.
- Seashore, C.E. (1919, 1939, 1960). Measures of musical talents (original version 1919, revised 1939, latest edition 1960). New York: The Psychological Corporation.
- Seashore, C.E. (1938). Psychology of music. New York: McGraw-Hill.
- Serafine, M.L. (1979). A measure of meter conservation in music, based on Piaget's theory. Genetic Psychology Monographs, 99, 185-229.
- Serafine, M.L. (1980). Piagetian research in music. Bulletin of the Council for Research in Music Education, 62, 1-21.
- Sims, W.L., Moore, R.S., & Kuhn, T.L. (1982). Effects of female and male vocal stimuli, tonal pattern length, and age on vocal pitch-matching abilities of young children from England and the United States. Psychology of Music, Special Issue: Proceedings of the Ninth International Seminar on Research in Music Education, 104-108.
- Sluckin, W., Hargreaves, D.J., & Colman, A.M. (1983). Novelty and human aesthetic preferences. Chapter 11 in J. Archer and L. Birke (eds.), Exploration in animals and humans. Van Nostrand Reinhold.
- Sopchak, A.L. (1955). Individual differences in responses to different types of music in relation to sex, mood, and other variables. Psychological Monographs, 69 (11).
- Spelt, D.K. (1948). The conditioning of the human foetus in utero. Journal of Experimental Psychology, 38, 375-376.

- Stirnimann, E. (1940). Psychologie des neugeborenes Kindes. Zurich: Leipzig.
- Stuart, M.H. (1981). Generating qualitative and quantitative changes in music teaching skills. Psychology of Music, 9 (2), 3-16.
- Trehub, S.E., & Rabinovitch, M.S. (1972). Auditory-linguistic sensitivity in early infancy. Developmental Psychology, 6, 74-77.
- Tversky, A., & Gati, I. (1978). Studies of similarity. In E. Rosch and B.B. Lloyd (eds.), Cognition and categorization (pp. 79-98). London: John Wiley & Sons.
- Walker, E.L. (1973). Psychological complexity as a basis for a theory of motivation and choice. In D. Levine (ed.), Nebraska symposium on motivation 1964. Lincoln: University of Nebraska Press.
- Watkins, J.G., & Farnum, S.E. (1954). The Watkins-Farnum Performance Scale. Winona, Minn.: Leonard Music.
- Wiebe, G. (1940). The effect of radio plugging on students' opinions of popular songs. Journal of Applied Psychology, 24, 721-727.
- Wieck, K.E., Gillfillian, D.P., & Keith, T.A. (1973). The effect of composer credibility on orchestra performance. Sociometry, 36, 435-462.
- Wing, H.D. (1948). Standardised tests of musical intelligence. National Foundation for Educational Research.
- Wing, H.D. (1968). Tests of musical ability and appreciation (second edition). London: Cambridge University Press.
- Zimmerman, M.P. (1975). Research in music education with very

young children. Report of the Fourth International Seminar on  
Research in Music Education. Wellington: New Zealand Council  
for Educational Research.



## Supplementary Appendices

### 1. Generality of findings

It should be pointed out that in the research described in this thesis, as in much cross-sectional developmental research, the extent to which findings may be generalised is open to debate. This is because age cannot be manipulated in the way that most true independent variables can, and therefore problems arise in selecting different age groups: it is, strictly speaking, impossible to control for all variables other than age which might influence the results. A typical problem might be that arising in a study where the groups being compared attend different schools because of the difference in their ages, in which case the effects of school environment are difficult to separate from those of age. The same can apply to many environmental considerations which vary systematically with age for the specific sample under investigation but not for the general population. In all developmental research, age as an "independent variable" is never really varied independently, because it is inevitably confounded with cohort effects.

## Supplementary Appendices

### 2. Non-independence of results: Experiments 1 & 2

It was argued on page 80 that in order to throw more light on the somewhat anomalous developmental trend found in Experiment 1, it would be interesting to investigate the responses of subjects both older and younger than those involved in that experiment. This was undertaken in Experiment 2. However, it must be stressed that because the data for the 8- to 9- and 10- to 11-year-olds in the latter study were a subset of those in the former, the results of the two experiments did not arise from two completely separate sets of data. Although the labelling of "Experiments 1 and 2" might be considered slightly misleading, no errors will arise so long as the results of the two experiments are not interpreted as if they were independent.