CHAPTER 5

THE FIRST STUDY

Chapter Overview

This chapter presents the first psychometric study. Two intact Years 5-6 at two government primary schools (N = 151) were administered three psychometric batteries:

1. marker tests for a three-factor operationalisation of the Luria model (Fitzgerald, 1978), together with a new test, the *Pitch Contour Inversion Test*, hypothesised to act as a marker test for simultaneous synthesis in the musical domain;
2. the *Music Evaluation Kit* (MEK) Part I *Pitch Discrimination* and Part V *Patterns Recognition* (Bryce, 1979); and,
3. an original instrument, *Semantic Evaluation of Spectral Density*, developed to replicate the study of Voss (1975) with the greater sensitivity afforded by a semantic differential design.

The results of the multivariate analyses supported Voss and Clarke (1975, 1978) that fractal music is preferred to other musics generated from random Brownian motion algorithms, and indicated that successive synthesis was important in this task. The relationships between performance on the MEK and the Luria model information processing dimensions were somewhat more equivocal. The low success rate on the *Pitch Contour Inversion Test* suggested that this task was too difficult for most subjects. This suggestion was examined by a reappraisal of the task in terms of the Luria model.

Chapter 5 has eleven sections. *Section 5.1* discusses the formulation of the research questions. *Section 5.2* describes the MEK as an instrument to measure musical ability. The third section describes the development of an instrument to measure perception of musical contour as a potential musical marker test for the Luria model battery. *Section 5.4* describes the development of an instrument to replicate the study of Voss. *Section 5.5* overviews the battery of tests for the three-factor operationalisation of the Luria model. The experimental situation is described in *Section 5.6*, and some aspects of the internal and external validity of the experimental situation are discussed in *Section 5.7.* In *Section 5.8* the research questions posed earlier are formulated into seven experimental hypotheses. *Section 5.9* presents the data analysis. *Section 5.10* interprets the results in terms of the Luria model, and the final section makes recommendations for the second study.

5.1 Formulation of the research questions

A number of important issues, as well as suggestions for further investigation, emerge from the review of the literature in Chapters 2, 3 and 4. First there is the claim for generalisability of the operationalised Luria model in the music domain. While much of the music cognition research evidence can be classified into a Lurian framework, this particular cognitive model has not been used previously to investigate individual differences in music ability. Such an endeavour faces several problems. To begin with, the literature is somewhat ambivalent about hemispheric specialisation for music processing. Luria was not. He argues for right or non-dominant hemispheric involvement in music perception. Whereas the operationalised model does not require cerebral localisation of the three (or four) processing dimensions measured, it must be noted that the findings of Luria, on which the marker tests are based, were mostly concerned with lesions to the left hemisphere. The first research question is, then: To what extent does the present operationalised Luria model account for variability in music information processing?

A subsidiary question follows: Do we need musical equivalent tests of Luria dimensions, or can variance in musical perception be adequately accounted for by relationships between existing marker tests and more standard measures of musical ability? Parallel tests might provide an answer to the question of whether musical information is processed in a similar manner to non-musical information. Luria's position that speech is not processed similarly to music, even though both share an auditory modality, should be noted. Certainly the outcome to Research Question 1 may set some constraint on the generalisability of the Luria model if music is not processed in a similar manner to other information. If music is processed in a similar manner then the scores on musical analogues of marker tests should load on to the same components as those original tests. To summarise as the second research question: Can a reliable musical test be developed as an analogue to any of the marker tests already employed in the operationalised Luria model, assuming that such a musical test is required at all?

The issue on which the literature is in considerable disagreement is the extent to which variance in musical ability can be attributed to musical aptitude or "giftedness" as distinct from the effects of musical training or experience. If, for example, learning in music parallels language acquisition, then to what extent is musical aptitude an identifiable trait, and could such a trait be explained by the information processing dimensions of simultaneous, successive and executive syntheses? Correlations of musical ability with successive and simultaneous syntheses could indicate the extent to which musical information is processed like other stimuli such as visual imagery or speech, or alternately, the extent to which successive and simultaneous syntheses operate within the musical domain. Specific relationships of music aptitude with executive synthesis could inform the various interpretations of attention and the role of enculturation in music learning.

Thus the third research question: Do children who show high measures for musical aptitude, or show high musical ability, have a distinctive pattern of cognitive abilities as determined by the operationalised Luria model?

A complementary issue from the literature review [Chapter 4] was that of the form that information in music might take. A number of contributors over the years have used an information processing framework, after Gibson, to conclude that our sense of musical meaning through expectancy is a function of the form of the musical stimulus, and that form is Markov or fractal. However, evidence in support of the study by Voss, such as Boon *et al* (1990), is strongly countered in studies by Nettheim (1992) and Gilden *et al* (1993), and alternate conceptualisations by Jackendoff (1991) and Narmour (1990). This suggests a need to replicate Voss's study. Moreover, the subjects in the original Voss study were all adults. Whether children are sensitive to differences in artificial music with different spectral density functions, viz., $\beta = 1, 0$ or 2, has not been established. Winner, Rosenblatt, Windmueller, Davidson and Gardner (1986) suggest that children do not attend to the aesthetic properties of music in the same manner as adult listeners; rather, aesthetic perception is multidimensional and develops with age. The perceptual requirements of children for novelty in order to maintain interest could also be of a dissimilar level to adults, especially given the quicker conceptual tempo of children, their more limited exposure to musical experience, and the possibility that the tertiary zone of the third functional unit may yet to have reached complete maturity. The fourth research question addresses a replication of Voss's study but with younger subjects: Do children prefer algorithmically generated fractal music, spectral density function slope $\beta = 1$, to white music ($\beta = 0$) or brown music ($\beta = 2$)? And as musical sensitivity is a frequently observed characteristic of music ability, research question five asks whether any such preferences for fractal music are stronger for children with high musical ability.

5.2 *The Music Evaluation Kit test battery*

Based on the review of the literature on music aptitude testing in Chapter 3 it was decided to use the *Music Evaluation Kit* (MEK), developed by the Australian Council for Education Research and the University of Melbourne (Bryce, 1979), to measure musical ability in the area of pitch discrimination [Appendix A]. The MEK is a battery of "domain-referenced" music tests which measure skills acquisition in seven areas of music learning: discrimination in pitch, sound duration, volume and tone colour, recognition of patterns, instrument identification, and knowledge of musical symbols. Profiles of student scores provide the teacher with "information about individual strengths and weaknesses" (p. iii). Parts of the MEK, then, may in principle be suitable for determining individual differences in those aspects of musical ability which predict aptness for future learning (Carroll, 1994). Since "the kit is not a selection or grading instrument, but a teaching aid which may be used for diagnostic work" (p. iii), such use is not in conflict with the original aims of the instrument.

The seven areas for testing are based on earlier batteries reported in the literature, in particular, Seashore's revised battery for musical sensitivity, with sub-tests for pitch, volume, time, timbre, rhythm and tonal memory. Each test in the MEK consists of twenty items presented on audio-tape which require a multiple(four)-choice response on a printed answer sheet. The authors claim that the sample of twenty is representative of the domain in question. The MEK does not include aesthetic judgements as these have problems for test reliability.

The most suitable MEK test for this study is *Part V: Pattern Recognition*. It would seem particularly suitable for measuring ability at temporal substructuring. However, since "it is generally assumed that pitch discrimination is basic to playing and listening to all forms of music" (Bryce, 1979: 2), *Part I: Pitch Discrimination* was also included to provide baseline measures of individual differences in those aspects of musical abilities which are specifically related to pitch.

The MEK is described as criterion-referenced in contrast to norm-referenced music tests such as Gordon's MAP (Gordon, 1979). However, the case for criterion-reference is somewhat undermined by the author's interpretation of outcomes. Bryce notes that a high score (18-20) may indicate mastery of a domain, whereas a score in the range 12-17 indicates "some reinforcement" required, and "students scoring below 12 may require ... extensive help with the skills and concepts that have not been understood" (p. 8). To then suggest, as Bryce does, that these scores cannot be used to rank students seems unreasonable, albeit pedagogically commendable. The complete battery of MEK tests was extensively trialled by Bryce across school systems and states, with age cohorts two to three years older than the mean age of the intended subjects. High levels of correct response (> 85%) were interpreted as indicating an item on which criterion could be met by musically able younger subjects. It is not clear, then, why the author expresses puzzlement over low indices for discrimination on those items with such high levels of correct response. Another possible limitation is that most of the items in the MEK Part I do not first establish a keyality. Radocy and Boyle (1988) argue that discrimination between two tones without reference to keyality does not require musical thought. In Part V some of the patterns are of only three notes, less than the six recommended by Karma (1985) for use in testing with children.

Nevertheless, there are also several advantages in using this instrument. The MEK is designed for administration on high school entry, and so should be somewhat more rigorous for subjects of a similar age in Year 6 and a year younger in Year 5. For the purposes of research in Australian schools, the MEK has the advantage of its verbal instructions being delivered with an Australian pronunciation on the test tapes, over tests where the instructors have American accents. Moreover, there has been no other test of a similar nature produced since 1979 in Australia (Bryce, 1993).

5.3 Instrumentation to measure the perception of musical contour inversion

Chapter 3 reviewed the cognitive similarities of perception of spatial and auditory contour (e.g., Terhardt, 1988). A good account of the variance in ability at spatial reasoning can be given by measures on simultaneous synthesis. Marker tests include: *Fitzgerald Matrices A*, in which subjects have to reproduce a closed geometric shape drawn on a 3 x 3 dot matrix after a brief exposure; *Fitzgerald Matrices B*, similar to *Matrices A* but requiring the reproduced shape to be rotated through 180 degrees; and the *Paper Folding Test*, which requires the mental unfolding of a marked sheet of paper. It was decided to design an instrument which would function as an auditory analogue to one of the marker tests for simultaneous synthesis. The *Fitzgerald Matrices B* test was chosen because it afforded manipulation of the stimulus contour.

The *Pitch Contour Inversion Test* [Appendix B] draws on Dowling (1971), who investigated the recognition of brief melodies after transposition and inversion. Two forms of inversion were used, contour preserving and exact interval inversion. The rationale for Dowling's study was that "many problems in visual pattern recognition find direct analogues in auditory pattern recognition. Two such issues are: what distortions of detail leave the pattern recognisable, and what are the effects of various changes of orientation of shape on recognition?" (p. 348). Dowling found that recognition of inversions was better than chance, but was significantly weaker than recognition of repetitions and transpositions. Dowling concludes, as does much of the related literature [Chapter 3], that "contour and interval size are handled in different and largely independent ways in cognitive processing" (p. 349).

Krumhansl, Sandell and Sergeant (1987) reviewed an earlier study by Balch (1981) which used melodic inversion, retrograde and retrograde inversion to test perceived relatedness of original melodies to these variations. Ratings were highest for inversion and retrograde. Krumhansl *et al* then investigated the perception of tonal hierarchies in tone rows used in serial music by employing the same "mirror imaging" as Balch. The results suggest that listeners can perceive invariance across such variations, but that listeners with musical experience in such a genre perform much better than those without. It could be predicted, then, that the ability to detect melodic contour inversion in younger subjects might show some relationship to musical experience or measures of musical aptitude, and possibly with scores on either simultaneous or successive synthesis.

The *Pitch Contour Inversion Test* [Appendix B] required subjects to select the inversion of a probe musical phrase from a set of four lure phrases. The stimuli were two-bar monophonic phrases of crotchets in common time, ranging from 3 to 15 tones in duration, performed with a digital piano timbre within the compass of the bass and treble clefs plus an octave above. The phrases were similar to those used in aural tests of the Australian Music Examinations Board (AMEB) Grade examinations. The phrases were generated on an Apple IIGS computer with the program *Music Studio*, recommended by Webster (1989) as appropriate composition software for research with children. The phrases were recorded with instructions on to cassette tape, and played back on the subjects' school assembly hall cassette tape player. The inversion targets were made by replacing each note of the probe with its exact tonal inversion, e.g., an ascending perfect 4th in the probe was replaced with a descending perfect 4th in the target; a descending minor 7th in the probe was replaced by an ascending minor 7th in the target. Thus keyality was not strictly preserved with phrases in minor keys. Distractor lures included transpositions of the stimulus phrase up or down an octave, the phrase played in retrograde, and the phrase repeated unchanged. The probe phrase was played twice, followed by the lures. The order of the different types of distractors was determined by random number generation. Subjects marked on an answer sheet "1", "2", "3" or "4" for which lure they thought was the inversion. The instructions emphasised listening for the piece played "upside-down". Several practice items with feedback were undertaken immediately prior to testing.

A bank of 20 items was trailed with forty Year 6 children who were not subjects in the first study. Results indicated that this task was very difficult for most subjects. Alpha reliability coefficients failed to reach a satisfactory level. Only four items were scored in the range of 40% to 60% correct. It was noted that two of these items had the longest strings of eight crotchets, whereas items scoring apparently at chance were typically shorter and slower. This suggested that a phrase may need to be of a sufficient length and tempo for a subject to attend to the contour rather than the individual intervals. Also, three of these items began with a rising interval, as do nearly 80% of all melodies in Western music (Pont, 1990). These four successful items were retained. Eleven items with similar contour type and length were added to make a set of fifteen for the *Pitch Contour Inversion Test* instrument used in the first study. It was also noted that the most attractive distractor was the retrograde. The instructions were revised to emphasise the "upside-down" nature of inversion, and the practice period was extended.

5.4 Instrumentation to replicate the Voss study

Voss (1975) [Chapter 4] investigated aesthetic responses to computer generated music with spectral density functions of $1/f^0$ - white noise or white music, $1/f^2$ - brown noise or brown music, and $1/f^1$ - pink noise or fractal music. Subjects rated fractal music over the others as being most like real music.

A programmable method for the generation of $1/f^\beta$ auditory sequences is random midpoint displacement (Voss, 1988) or successive random additions (Bunde & Havlin, 1994). A self-affine function Z(x) is generated in the interval $0 < x < 1$ through the iterative addition of random values Z at locations $x = (1/2)^n$.

At n=1, x=1/2,

$$Z(0) = Z(0) + R_a$$

$$Z(1) = Z(1) + R_b$$

$$Z(1/2) = [Z(0) + Z(1)]/2 + R_c \qquad (5.1)$$

where R_i are random displacements with variance $\sigma_1 = (1/2)^H$ and H is the Hurst exponent [Chapter 4]. In general, the nth iteration adds random values R_i to all Z values with variance $\sigma v = (1/2)^{nH}$. Based on a suggestion of Voss reported in Gardner (1978: 25), Landini (1992) employed this algorithm to generate examples of fractal music using a QBASIC program [Appendix C]. Each "1" in a series of binary numbers acts as a trigger for a random number generator (dice) whose outputs are added to produce each value in the time series. Each nth column (counting from the right) of the generated binary chart could be made to have a strict self-similarity with the n = 1 column (or any n = k column) by sampling down the column in lots of 2^n integers and recoding each lot as "1" or "0" depending on the contents. The triggers exhibit self-similarity over successive re-scalings by 2^n. Such a cascade of dice-triggers is a Markov process, i.e., future values depend on the present value only and not on previous values at t-1, t-2, etc. It is not immediately obvious that this process will necessarily produce a fractal output time series from the random values given by the dice. However, as Gardner points out, the series produced by this multiplicative addition does share a property with fractal time series, viz., local effects (the right column) are more significant in determining values than more global effects (left columns), rather like an amplitude-modulated waveform.

To demonstrate the extent to which the output of Landini's program is 1/f-like, Landini provides a plot of the autocorrelation over time which showed a decay to zero in about 70 seconds. This was checked by an examination of the autocorrelation function over a run of 256 new output values. The lag period was set at 24. The first zero crossing occurred after the third lag = 72 outputs, which compares with Landini's plot if his abscissa unit is mislabelled as "seconds" rather than "outputs" [Appendix C]. Otherwise the autocorrelation function seems somewhat too truncated for true 1/f-like behaviour (Gregson, 1994). A Fourier analysis of the "dice" outputs of Landini's program with the number of "dice" = 5 showed a large zero peak and evident periodicy of 32 [Appendix C]. Such periodicity would seem to undermine any claim to purity of fractal form, not that Landini or Gardner, are making such a claim. Gardner does claim, however, that one cannot tell the difference between the output of the random multiple addition algorithm and a genuine 1/f signal. Landini also provides a log-log plot of the spectral density function of sequences from his algorithm. The general trend follows a 1/f curve, but the function plot itself displays considerable variance [Appendix C].

The more salient criterion, then, is how music-like the algorithm outputs are. There is no claim that the samples by Landini sound like real music. But Gardner does argue that 1/f music should show similar probabilities for successive intervals as real music, in contrast to brown music where small intervals are generated with too high a probability, and in contrast to white music where all intervals are equally likely. Thomas (1991) claims that such music is "generally regarded as accessible, interesting, pleasant, and containing a balance of predictability and unpredictability" (p. 100). This claim was tested in a simple manner, following Cohen, Trehub and Thorpe (1989), with interval frequency tables such as were gathered by Hanna (1965) on the piano works of Schubert. As the Landini algorithm generates only a single line of notes, comparison with multiphonic real music is problematic. To minimise this difficulty, examples of music were used which had a defined monophonic melodic line. The examples selected were:

Bach: 1. Flute Sonata Bb, 2nd mvt, flute part;
2. *Chorale*, Cantata No. 147, arr. guitar;
3. *Aria* from Orchestral Suite D, arr. organ, top treble line.

Debussy: 1. *Claire de Lune*, arr. flute and piano, flute part;
2. *Le Petit Berger*, arr. flute and piano, flute part;
3. *Syrinx*, solo flute.

Lennon and McCartney: *When I'm Sixty Four*, lead vocal line.

Pemberthy: *Six Miniatures* for flute/violin/piano, flute part.

Ascending and descending intervals were counted with rests ignored. The mean interval frequencies as percentages of the total intervals are reported in Appendix C. This survey was repeated for ten outputs of the algorithm. As Landini notes, the algorithm does not include note exclusion rules for "playing" in a particular key or mode. To impose note exclusion post hoc, intervals were collapsed by degree. For example, both major and minor thirds were counted as thirds, diminished and perfect fourths were both counted as fourths. The counts from the ten outputs of the algorithm were averaged. To compare the pitch occurrences of the various real and algorithmic music examples, the frequency tables were subjected to a contingency table analysis.

To avoid limitations of contingency analysis from low frequencies, intervals of a seventh and higher were excluded from the analysis. The results are summarised in Table 5.01.

p levels	**Bach 1**	**Bach 2**	**Bach 3**	**Debus 1**	**Debus 2**	**Debus 3**	**Beatles**	**Pembert**
white	.034	**.106**	.007	.0001	.004	**.171**	.035	.030
brown	.0001	.0003	.0001	.0001	.0001	.0001	.0001	.0001
1/f	.003	.002	.039	.0001	.008	.005	**.242**	.0008

Table 5.01 Contingency table analysis of differences between real melodies and fractal, brown and white musics

While acknowledging the simplicity of this analysis, it is interesting to note that in general the interval frequencies of the real melodies were significantly different from all the computer musics, including the 1/f series. Only the piece by *The Beatles* was not different from 1/f . These results would seem to support the criticism in Chapter 4 that the claim by Voss for music in general to be 1/f-like does not apply to any one piece in particular. An elaboration on such an interpretation is that compositional style, or use of motifs, is a more dominant effect in any particular case than a general statistical trend. Westhead and Smaill (1993) found significant differences in the use of characteristic motifs between different composers, Bach and Tchaikovsky, and even between different compositional forms by the same composer, such as the fugues and inventions of Bach. Perhaps surprisingly, Pemberthy's tone row composition was significantly different from white noise, while a composition of each of Bach and Debussy were not significantly different. This underscores the non-isomorphic relationship between real music and algorithmic generated tone series. Whereas algorithmic white music will contain an equal distribution of intervals, a composed score with a fairly equal distribution of intervals will not necessarily be white, i.e., sound random, while a piece that sounds random-like, may not contain an equal distribution of intervals. This point was emphasised by Nettheim (1992).

Overall, the limitations to the fractal characteristics of the Landini program possibly provided a more stringent perceptual task for subjects in the replication of the Voss study. Given the conflicting analyses of the autocorrelation function of Landini's outputs, the length of the stimulus samples were set to fifteen seconds, so that correlation was still moderately high for all lag times. Nettheim also comments on the difficulties of preserving 1/f spectral properties in music with multiple dimensions of 1/f-like behaviour. "It is clear that the resultant processes do not have the properties of their two constituents. For example, the resultant of 'white noise' pitches and 'white noise durations' is by no means 'white noise music' " (Nettheim, 1992: 137). Nevertheless, to ensure that the algorithmic fractal music was similar to the examples based on 1/f noise by Voss [Figure 4.05], which have durational distributions as well as pitch, the Landini algorithm was used to generate 1/f-like pitch series (median A4 treble clef) together with a 1/f -like distribution of durations (semi-quaver = 0.2 sec). Parallel QBASIC algorithms were written to generate contrasting white and brown pitch series [Appendix C].

These programs were used in a trial instrument, *Realness of Fractal Music.* Fifteen examples of music, of which twelve were computer generated, were presented for assessment on a five-point Likert scale from "Very much like real music" to "Not at all like real music". The computer generated pieces included: six examples of $1/f^{\beta}$ music with two each of $\beta = 0$, $\beta = 1$, and $\beta = 2$; three pieces generated by other chaotic algorithms (Pressing, 1987); and two pieces demonstrating perfect autocorrelation, i.e., exact repetitions (Pickover, 1986). Anticipating that the novelty of computer generated music may cause general responses to cluster towards the "Not at all like real music" end of the scale, three extracts from real music were included as a baseline. Two were by composers who claimed to be extrinsically influenced by natural images, Schulthorpe and Messiaen. The other was an extract from a solo 'cello partita of Bach whose music Voss and Clarke (1978) report as displaying 1/f behaviour.

The trial was conducted with students of two intact Years 6 at the two schools where the first study was undertaken. Some of the music examples were up to 60 seconds in duration. This was too long for most subjects at School A, so only the first 15 seconds of each piece was played at School B. Obviously, validity is compromised in combining scores from both schools, but actual attention span seemed to be similar in both cases. The mean score for fractal music (3.40) was lower (towards the "real music" end of the scale) than the score for white music (3.76) but not lower than the mean score for brown music (3.33). Moreover, the mean for the examples of real music (2.57) was only just on the positive side of the scale. Adjustment of the likeness scores for fractal, white and brown musics by the likeness of real music score showed no difference in the relative position of the fractal music.

The trial results highlight a problem with using real music examples in testing: the cultural embededness of aesthetic preferences for music. Minsky (1988) suggests that aesthetic development can be explained by psychological associations with pain and pleasure across all modalities in early childhood. Whatever the cause, idiomatic identification threatens to account for nearly all of the variance on any dichotomous scale such as 'like-dislike'. It was decided to address this issue by the use of semantic differential scales. Semantic differentials can be used to assess meaning within a semantic space defined by a set of scales, each scale being delineated by a bipolar pair, or a factored grouping of bipolar pairs of adjectives, e.g., "happy-sad", "black-white", "hard-soft" (Osgood, Suci & Tannebaum, 1957). The connotative meaning of a concept, then, is measured as a location in this multi-dimensional semantic space. The location has both distance and direction from the origin: distance is measured as position on the scale, direction is indicated by choice of polarity. These measures are interpreted as indicating intensity and quality of meaning respectively. The semantic differential is designed for stimuli from several modalities - visual, auditory, emotional and verbal. This approach is supported by Umemoto (1992) who suggests that "complex musical concepts with ambiguous meanings ... can be studied by the direct use of words" (p. 124).

In selecting bipolar adjective pairs the Hevner adjective list was considered. Eight clusters of adjectives are arranged around a "mood clock" (Hevner, 1936). However, Farnsworth (1953) showed that operationalising these groups did not describe internally consistent mood patterns. Consequently, it was decided to use bipolar pairs which had demonstrated reliability for general perceptual evaluation. Geake (1991) used a semantic differential instrument to evaluate affective responses to images of natural form. Four of the bipolar pairs used in that study were "kind-cruel", "fair-unfair", "white-black", and "beautiful-ugly". A fifth pair, "fresh-stale", was added. Subjects rated these on a five step scale, presented in random order and polarity, for each of twelve music items [Appendix D]. For this *Semantic Evaluation of Spectral Density Test* instrument, it was decided to control for music experience by only using algorithmically generated examples, viz., the fractal, white and brown algorithmic music generated by the modified 1/f algorithm. By avoiding both classical and popular genres, musical cultural context was restricted but still within the broad genre of diatonic scaling. The synthetic computer timbre was assumed to be familiar to all subjects.

5.5 The Luria model battery

The operationalised Luria model battery employed in the first study was the three factor model with simultaneous, successive and executive synthesis [Appendix E]. These and similar instruments have been used in a number of previous studies by Fitzgerald and colleagues. The strengths and limitations of this battery were reviewed in Chapter 2.

Simultaneous synthesis:

Fitzgerald Matrices A

This is a recall task of line figures. Twenty different figures were drawn by joining some of the dots on a 3 x 3 dot matrix to make a closed figure. Slides of the figures are each shown to the subjects for five seconds. This is immediately followed by ten seconds response time for subjects to attempt to reproduce the figure on a 3 x 3 dot grid on an answer sheet. Items are scored: 1 all correct, 0.5 for > 60% correct, 0 otherwise. Total scores are computed for analysis.

Fitzgerald Matrices B

This task is identical to *Matrices A*, except that subjects are instructed to reproduce each figure as it would appear if it were rotated to be upside down. The instrument is scored the same as for *Matrices A*.

Paper Folding Test

Figures on the answer page indicate how a sheet of paper had been folded two or three times and then had a hole or holes punched through the folded portion. The task is to choose which of five lures represents the pattern of holes of the unfolded paper. There are 10 items which are scored 1 mark each for a correct choice.

Successive synthesis:

Auditory Word Span

This task tests ability to recall sequences of words. A set of words is read aloud once. Subjects are instructed to write down the set in the exact order in which the words were read. There are twenty sets from three to nine words long. Each correctly sequenced word scores 1. Item scores are totalled for analysis.

Auditory Number Span

This task is identical to the *Auditory Word Span* except that sets of numbers are used. The sets are from four to twelve numbers long.

Executive synthesis:

Letter Search

This task tests speed and accuracy of selecting vowels from a string of mixed letters. Sixteen strings of 24 letters are used. Subjects have five seconds for each string. Each selected vowel scores 1; each selected consonant scores -1. Item scores are totalled for analysis.

Odd/Even Number Search

This task is identical to the *Letter Search* except that strings of numbers are used. Subjects are instructed to select either the odd or the even numbers. Each correctly selected number scores 1. Each incorrectly selected number scores -1.

Letter/Number Attention Span

This task is identical to the *Auditory Number Span* except that mixed sets of numbers and letters are used. Subjects have to write either the numbers or the letters, as indicated before the set is read out. The sets are from five to ten characters long.

5.6 Experimental situation

The subjects (N = 151) were 10 - 11 year olds in Years 5 and 6 from two North Coast government primary schools. Permission to conduct research in NSW State Government schools was obtained from the North Coast Regional Office of the Department of School Education. Testing was undertaken over two sessions at each school. Session 1 involved the administration of the Luria model battery and the *Pitch Contour Inversion Test.* The MEK Parts I and V and the *Semantic Evaluation of Spectral Density Test* were administered in Session 2. There was a total of 24 instances of absenteeism across the two sessions in both schools. This appeared unsystematic except for Session 1 at School B where six subjects were attending a Gifted Writers Camp. These subjects were subsequently tested at Session 2. Session 1 was conducted with each Year as a whole in the school assembly hall. Session 2 was held in individual classrooms. Although the item trials had been conducted with Year 6, Year 5 were selected for Session 1 to allow for the Session 2 follow-up with the same subjects during the following year. There was a subject mortality of 15 between testing periods of October and April. A random group of 18 subjects was re-tested with the complete Luria model battery to check for any maturation effect between sessions. Marker test scores of these subjects in the October and April sessions were compared using paired t-tests: none of the differences were significant.

5.7 Experimental validity

The experimental situation was typical of situations in other psychometric studies in terms of commonly accepted criteria for research design which contribute to internal and external validity, viz., the adequacy of the design and the control of extraneous variables. The preceding sections concerned with instrument design have considered the central issue of validity, the extent to which the variables represent the concepts planned to be tested. Internal validity is defined as "the validity with which statements can be made about whether there is a causal relationship from one variable to another" (Cook & Campbell, 1979: 38), i.e., the extent to which the experimental results are not due to extraneous variables. Such extraneous variables can include: significant events; maturation of subjects; increased familiarisation with testing; changes in instrumentation; diffusion of treatment effects between groups; and, compensatory responses by and/or for subjects in less desirable treatment groups. Other extraneous variables can arise from interactions with the selection of experimental groups. Validity can also be compromised through statistical regression. External validity is defined as "the approximate validity with which conclusions are drawn about the generalizability of a causal relationship to and across populations of persons, settings and times" (Cook & Campbell, 1979: 39). Threats to external validity can arise from interactions with selection and setting.

The number of subjects, 151, is in excess of the sample size recommended by Stevens (1986) to minimise Type II error in MANOVA designs with small effect size. To undertake this research in schools, intact classes had to be used. As tests were administered to either the whole school or to concurrent classes, and all subjects were administered the same battery, diffusion between experimental groups was not an issue. The schools were selected for internal diversity of socio-economic profile in order to maximise the generalizability of the results. School A was located in a rural village with a rural feeder area; School B was located in a rural city with a feeder area which included middle and lower socio-economic strata.

Age-stage differences were partly controlled by using subjects from the age band (10 - 12 years) of Years 5 and 6. According to Golden (1987) [Chapter 2], this age band falls just after the onset of maturity of the tertiary zone of the third functional unit making it an optimal age for sensitivity to the Luria model battery. Radocy and Boyle (1988) suggest that overall harmonic awareness is stable by age six to eight years, and Gordon (1993) states that ability to audiate is stable by age nine. Thus this age range should also be suitable for s sensitive response to the MEK tests. The adequacy of the language used for the test items and the instructions was confirmed by the children used as subjects for the trialing of the instruments, and the classroom teachers of subjects participating in the study. There were no severely hearing impaired children among the subjects. Each battery was administered by the same person at both sites and only two testers were employed throughout the three sessions. There were no cases of tests being incomplete due to misunderstanding of instructions. Statistical regression towards the mean was not a problem here because selection was not based on any predetermined level of competence at any of the tests. The reliability of the test instruments was assessed during the data analysis to ensure that it was adequate in the experimental situation.

Two schools were used to partially control some of the many extraneous variables which arose. These included: organised absences such as school camps during the testing sessions; unforeseen absences due to contagious illnesses; access to music lessons outside of school; strength of the school's music program; personality of the teacher; teacher attitude towards research; and parents' interest. Testing did not contribute to an increased school work load as it was undertaken during normal lesson times. All participating principals and teachers expressed enthusiasm for the project. Teacher attitude was seen as an important contribution to the likelihood of all subjects participating genuinely in the testing. Parental approval was sought for children's participation. Only one child was withdrawn because of parental non-approval.

There was a question of how to control for differences in the novelty of computer generated music. However, it was readily apparent that all subjects were very familiar with computers in their school context and thus also familiar with the timbre of a micro-computer sound card. All participating classrooms had computers which the children frequently used for a variety of class and recreational activities. This approach could not of itself control the interactive effect of familiarisation with computer music outside of the classroom. It was intended that the choice of participating schools across a range of socio-economic strata would provide some degree of control for this. Anecdotal comments from the subjects, however, indicated that the use of home computers was similarly substantial across all participating classes. Further questioning revealed that fractal-type music was novel for all subjects. The subjects commented on the novelty of all the test instruments used. It seems unlikely, then, that any effect of familiarisation with the style of this testing would be increased by school testing unassociated with this research.

Subject mortality was not so high to be a problem with 91% retained. Although testing took place over several days, to fit in with school programs and to reduce test fatigue, occasional absences reduced the number of subjects who completed all tests in Session 1. Classes were not affected by varying membership or changes in group dynamic. Certainly the group dynamic was different for each class, but the use of five classes in the sample was intended to provide some control for this.

It should be noted that the extraneous variables, such as they were, conspired to reduce the effect size rather than to produce a spurious result. Thus it was the adequacy of the testing outcomes, more than internal validity, that was threatened by the effects of these extraneous variables.

5.8 Experimental hypotheses

The Research Questions in *Section 5.1* were formulated into seven Experimental Hypotheses.

Hypothesis 1.1: That subjects who attain criterion on the MEK Parts I or V will have high measures of (a) simultaneous synthesis, and (b) executive syntheses.

Hypothesis 1.2: That there is a positive correlation between scores on the *Pitch Contour Inversion Test* and measures of simultaneous synthesis.

Hypothesis 1.3: That subjects who attain criterion on the MEK Parts I or V will have higher scores on the *Pitch Contour Inversion Test* than other subjects.

Hypothesis 1.4: That on the *Semantic Evaluation of Spectral Density Test*, subjects will rate fractal music more positively than white music and brown music.

Hypothesis 1.5: That on the *Semantic Evaluation of Spectral Density Test*, subjects with high measures of simultaneous synthesis will (a) rate fractal music more positively, and (b) rate white music and brown music less positively, than other subjects.

Hypothesis 1.6: That on the *Semantic Evaluation of Spectral Density Test*, subjects with high measures of successive synthesis will (a) rate fractal music more positively, and (b) rate white music and brown music less positively, than other subjects.

Hypothesis 1.7: That on the *Semantic Evaluation of Spectral Density Test*, subjects who attain criterion on the MEK Parts I or V will (a) rate fractal music more positively, and (b) rate white music and brown music less positively, than other subjects.

5.9 Data analysis

All data were analysed with SPSS - PC (Tabachnick & Fidell, 1983) [Appendix F]. Significances are reported at the 95% confidence level except where indicated otherwise. ANOVA procedures were used to investigate effects on several criterion variables individually in Hypothesis 1.3; MANOVA procedures were used to investigate the effects of some variables as a related set in Hypotheses 1.1, 1.2, 1.3 1.5, 1.6, and 1.7 (Bray & Maxwell, 1982).

Preliminary analysis

The six Luria model marker tests were reduced to three dimensions by a principal components analysis with Varimax rotation [Appendix F]. The components were interpretable in terms of the Luria model [Table 5.02]. The loading of the three marker tests for simultaneous synthesis, *Fitzgerald Matrices A* (MATRIXA) (.85), *Fitzgerald Matrices B* (MATRIXB) (.71), and the *Paper Folding Test* (PAPERF) (.68) on Component 1 indicates that this component reflects simultaneous synthesis. Similarly, Component 2, with loadings of *Number Span* (NUMBSPAN) (.83), *Word Span* (WORDSPAN) (.83) and *Letter/ Number Attention Span* (LETTNUM) (.49) reflects successive synthesis, and Component 3, with loadings of *Letter Search* (LETTSCH) (.76), *Number Search* (NUMBSCH) (.84) and LETTNUM (.39) reflects executive synthesis. LETTNUM loaded about equally on to Component 2 and Component 3, indicating the use of both successive and executive synthesis on this task.

Bartlett factor scores (sim, suc and exec) were computed to be used in further analysis, and also

	Component 1	Component 2	Component 3
MATRIXA	**.86**	.17	.10
MATRIXB	**.71**	-.04	.26
PAPERF	**.68**	.16	-.13
NUMBSPAN	.15	**.83**	.14
WORDSPAN	.02	**.83**	.09
LETTSCH	.04	.18	**.76**
NUMBSCH	.10	.13	**.84**
LETTNUM	.06	**.49**	**.39**

Table 5.02 Component structure of Luria model marker tests for Study 1

to partition subjects into three groups, high (sim1, suc1, exec1), medium (sim2, suc2, exec2) and low (sim3, suc3, exec3) on the three dimensions.

The normality of distribution of the experimental variables MEK1, MEK5, MUSISCR, FRACSCR, BROWNSCR and WHITESCR was checked by computing the coefficients of kurtosis and skewness. All but two variables had acceptable values of skewness and kurtosis. The distribution of WHITESCR showed excessive positive skewness, coefficient of skewness = 1.12, indicating a preponderance of low values. The distribution of MEK5 showed excessive negative skewness, coefficient of skewness = - 0.825, indicating a preponderance of high values. The distribution of MEK5 was also slightly leptokurtic, coefficient of kurtosis = 4.17, indicating that data distribution was somewhat peaked. However, these modest deviations from normality by only two variables should not present a compromise to statistical power (Stevens, 1986).

The possibility of differences between the subjects from the two schools was investigated by an ANOVA with SCHOOL as the independent variable and sim, suc and exec, and MEK1, MEK5, MUSISCR, FRACSCR, BROWNSCR and WHITESCR as dependent variables. Neither the multivariate effect nor any of the univariate effects were significant except for WHITESCR. Possible explanations for this effect are discussed below.

Hypothesis 1.1: That subjects who attain criterion on the MEK Parts I or V will have high measures of (a) simultaneous synthesis, and (b) executive syntheses.

The recommended MEK criterion level is 18/20 on each test (Bryce, 1979). Eleven subjects attained criterion on the MEK Part I (MEKIHI), fourteen subjects attained criterion on the MEK Part V (MEKVHI), and seven subjects attained criterion on both of the MEK Parts I and V. Successful performance on the MEK was determined by the attainment of criterion on either of the Parts I and V (MUSAPT), and was interpreted as indicating superior music ability. Nineteen subjects were thus classified as musically-able compared with their age peers. This proportion of the sample is similar to that found by Andreasen and Geake (1995) in a study involving Year 7 students (N ≈ 200) with the full MEK battery in which eighteen subjects attained criterion on at least two of the seven Parts.

The experimental hypothesis was tested with a 2-group MANOVA with MUSAPT as the independent variable and sim, suc and exec as dependent variables [Appendix F]. None of the effects were significant. The possibility that this null result was an artefact of small or unequal cell sizes was considered. For multivariate analysis, Tabachnick and Fidell (1983) recommend 20 data per cell to assure multivariate normality. Asmus and Radocy (1992) suggest ten as a minimum cell size for factorial experiments. The size of the MUSAPT group was increased by employing the more relaxed criterion of 17/20 for either of the MEK Parts. Given that the subjects in this study were about one year younger than the recommended age for administration of the battery (Year 7), this seemed justifiable. The size of the musically-able group rose to 37. The MANOVA was repeated. None of the effects, however, were significant. Inspection of the adjusted means showed that relaxing the criterion level resulted in a decrease in the mean scores of the musically-able group on sim (.23 -> .09) and exec (.32 -> .08) to near the sample means of zero . This result could support the claim by Bryce (1979) that the MEK is a criterion reference instrument, and that the appropriate criterion is a score of 18/20.

The failure to reach significance suggests that the null hypothesis associated with Hypotheses 1.1 should not be rejected with respect to (a) or (b) . A possible explanation is that variance in abilities on perception of pitch differences and pitch patterns is largely the outcome of music education experiences rather than cognitive abilities, although some individuals may be cognitively advantaged in such experiences.

Consequently, the suggestion made in *Section 5.2,* that the MEK scores could represent a continuous rather than criterion variable, was explored. First, it was conjectured that there should be a positive relationship between the scores on the MEK Parts I and V. A one-tailed Pearson's correlation between the scores on the two parts, MEK1 and MEK5, was significant (r = .30, p = .001). The two sets of scores were therefore added for a total MEK score (MEKTOT). The distribution of MEKTOT was checked for normality. The coefficients of skewness = .013 and kurtosis = 3.32 showed a close approximation to a normal distribution.

Next, it was conjectured that there should be a positive relationship between abilities on the information processing dimensions and MEK total scores. One-tailed correlations between MEKTOT and sim, suc and exec were significant for exec (r = .22, p = .016) and approached significance for sim (r = .16, p = .052). This result provides some tentative support for the experimental hypothesis that pitch discrimination involves simultaneous synthesis, and that attentional aspects are important in tasks of pitch and pattern discrimination. This analysis also suggests that the MEK scores may have some value as continuous variables in future studies of this research.

Hypothesis 1.2: That there is a positive correlation between scores on the *Pitch Contour Inversion Test* and measures of simultaneous synthesis.

A total score of correct responses on the *Pitch Contour Inversion Test* (MUSISCR) was computed. Responses to the individual items (MUS01-MUS15) were subjected to reliability analysis to determine whether reliability had improved over the unsatisfactory levels from the instrument trial [Appendix F]. Again, only four items were correctly answered at better than chance. However, the Kronbach alpha coefficient had risen to .595, significantly above zero (p

< .001), with each item contributing around 5% to the scale variance. These results suggest that, although this task is too difficult for many subjects, there is an acceptable level of homogeneity across the instrument. Consequently, analyses were undertaken with both MUSISCR and the MUS01-MUS15 responses.

To test the experimental hypothesis, a one-tailed Pearson correlation between MUSISCR and sim was computed. The correlation was significant (r = .18, p = .021). The null hypothesis associated with the Hypothesis 1.2 should be rejected.

To explore possible relationships, including interaction, between performance on the *Pitch Contour Inversion Test* and the three Luria model dimensions, a 3 x 3 x 3 MANOVA was undertaken with sim3, suc3 and exec3, the Luria model component scores partitioned into high-medium-low, as independent variables and the set of item results MUS01 to MUS15 as dependent variables [Appendix F]. For this exploratory analysis, significances are reported at the 90% confidence level and all univariate effects are noted. Significant multivariate effects are reported in Table 5.03. Significant univariate effects are reported in Table 5.04.

multivariate effect	Wilk's λ	significance	multivariate effect size
sim3 by suc3 by exec3	.215	p = .066	.175
sim3 by suc3	.432	p = .070	.189
exec3	.574	p = .011	.242

Table 5.03 Multivariate effects of 3-way partitions of Luria model components on *Pitch Contour Inversion Test* item scores

The multivariate effects in Table 5.03 for the three-way interaction could suggest that the three information processing dimensions of simultaneous, successive and executive synthesis all contribute to the information processing required on the *Pitch Contour Inversion Test.* However, inspection of the means [Appendix F] did not reveal any readily interpretable patterns. Table 5.04 shows that no factor, including the hypothesised dimension of simultaneous synthesis, related significantly to more than three dependent variables. Inspection of the *Pitch Contour*

Inversion Test items [Appendix B] for contour shape, whether hill, valley, slope or wave, and other aspects such as the number of changes in contour direction, the pitch range and the length, did not reveal any obvious confluence of pattern by factor.

	sim3	suc3	exec3	sim3 x suc3	sim3 x exec3	suc3 x exec3	sim3 x suc3 x exec3
MUS01			***				
MUS02							
MUS03				***			
MUS04			***				**
MUS05				*	*		
MUS06	**			**			
MUS07							
MUS08							**
MUS09							
MUS10	**						
MUS11						**	
MUS12		*	*			**	
MUS13							
MUS14				**			
MUS15							

* $p < .1$ ** $p < .05$ *** $p < .01$

Table 5.04 Univariate effects for 3-way partitions of Luria model dimensions on *Pitch Contour Inversion Test* items

The *Pitch Contour Inversion Test* was designed as a possible marker test for simultaneous synthesis in an auditory modality. The results above would suggest that the *Pitch Contour Inversion Test* is unsuitable for this task. This was confirmed using a principal components analysis with Varimax rotation of sim, suc, exec and MUSISCR, where Component 2 with a high loading of sim (.96) only attracted a low loading of MUSISCR (.33) [Appendix F].

Hypothesis 1.3: That subjects who attain criterion on the MEK Parts I or V will have higher scores on the *Pitch Contour Inversion* test than other subjects.

To test this hypothesis, a one-way ANOVA was undertaken with MUSAPT as the independent variable and MUSISCR as the dependent variable [Appendix F]. The MUSISCR effect was significant ($F(1,123) = 8.08$, $p = .005$, $\eta^2 = .062$). Inspection of the means [Table 5.05] showed that subjects who reached criterion on the MEK had higher total scores on the *Pitch Contour Inversion Test* (mean = 6.4) than subjects with below criterion MEK scores (mean = 3.48). The null hypothesis associated with Hypothesis 1.3 should be rejected.

	criterion on MEK I or V	below criterion on MEK I and V
MUSISCR	6.40	3.48

Table 5.05 Means of *Pitch Contour Inversion Test* scores for MEK criterion

As the effect size of 6% for the total score was modest, it was decided to further investigate this relationship by undertaking a 2-group MANOVA with MUSAPT as the independent variable and the set of item scores, MUS01 to MUS15, as the dependent variables [Appendix F]. The multivariate effect was significant (Wilk's $\lambda = .748$, $p = .001$, $\eta^2 = .252$). The univariate effects were significant on MUS07, MUS08, MUS09, MUS10, MUS12 and MUS13. Inspection of the means [Table 5.06] showed that subjects who reached criterion on the MEK had higher scores on these items of the *Pitch Contour Inversion Test.* Some 25% of the variance in success on these items could be accounted for by membership of the group which achieved criterion on the MEK Parts I or V. Whereas the *Pitch Contour Inversion Test* seems generally too difficult, six of the fifteen items were manageable by these musically-able subjects. Three of these six items, MUS08, MUS10 and MUS12, were also significant for the effects of simultaneous, successive and executive processing [Hypothesis 1.2]. Again, inspection of the contour shape and other attributes [Appendix B] did not reveal any features which distinguish these items from the others of this instrument.

	criterion on MEK I or V	below criterion on MEK I and V
MUS07	.80	.28
MUS08	.80	.23
MUS09	.80	.22
MUS10	.60	.16
MUS12	.80	.37
MUS13	.80	.19

Table 5.06 Means of *Pitch Contour Inversion Test* item scores for MEK criterion

Hypothesis 1.4: That on the *Semantic Evaluation of Spectral Density Test*, subjects will rate fractal music more positively than white music and brown music.

The scores on the five semantic differential scales of 'stale-fresh', 'unfair-fair', 'cruel-kind', 'black-white' and 'ugly-beautiful' of the *Semantic Evaluation of Spectral Density Test* instrument were summed for each of the three music types: fractal (FRACSCR), white (WHITESCR) and brown (BROWNSCR).

The experimental hypothesis was tested with two one-tailed paired t-tests between FRACSCR and WHITESCR, and between FRACSCR and BROWNSCR. In each case, the differences between the means of the combined scale scores were significant ($p < .0001$) [Table 5.07]. On the five-point scale, subjects rated fractal music (mean = 2.99) higher than white music (mean = 1.82), and higher than brown music (mean = 2.52). This evidence clearly supports the rejection of the null hypothesis associated with Hypothesis 1.4. The earlier findings of Voss and Clarke (1975, 1978) have been replicated here with children as subjects.

To further explore relationships between ratings of fractal, white and brown music across the five scales of this semantic differential instrument, the scores on the five semantic differential scales of 'stale-fresh', 'unfair-fair', 'cruel-kind', 'black-white' and 'ugly-beautiful' were computed

	mean on five-point scale
FRACSCR	2.99
WHITESCR	1.82
BROWNSCR	2.52

Table 5.07 Combined-scale means for ratings of fractal, white and brown music

separately for each of the three music types. The labels for these fifteen variables are listed in Appendix F. The differences in the mean scores on each scale between fractal and white, and between fractal and brown musics, were compared with a series of paired t-tests. In every case the difference was significant ($p < .001$) - the mean scale scores for fractal music were higher [Table 5.08]. Given the wide range of differences among the five semantic scales, this result further confirms that children prefer fractal form in auditory stimuli over the extremes of random or Brownian distributions of pitch fluctuations.

	fractal music	brown music	white music
stale-fresh	2.87	2.45	1.68
unfair-fair	3.14	2.64	1.91
cruel-kind	2.98	2.57	1.86
black-white	3.11	2.66	2.04
ugly-beautiful	2.90	2.29	1.67

Table 5.08 Scale means for ratings of fractal, white and brown music

Hypothesis 1.5: That on the *Semantic Evaluation of Spectral Density Test,* subjects with high measures of simultaneous synthesis will (a) rate fractal music more positively, and (b) rate white music and brown music less positively, than other subjects.

Hypothesis 1.6: That on the *Semantic Evaluation of Spectral Density Test,* subjects with high measures of successive synthesis will (a) rate fractal music more positively, and (b) rate white music and brown music less positively, than other subjects.

To test these two hypotheses, a series of two-tailed correlations were undertaken between the Luria model component scores sim, suc and exec, and FRACSCR, WHITESCR and BROWNSCR [Appendix F]. None of the correlations were significant [Table 5.09]. There is no evidence with which to reject the null hypotheses associated with Hypotheses 1.5 (a) and (b), and 1.6 (a) and (b).

	sim	suc	exec
FRACSCR	.02	-.04	-.05
BROWNSCR	-.006	-.03	.01
WHITESCR	.20	.05	-.09

Table 5.09 Correlations between Luria model component scores and *Semantic Evaluation of Spectral Density Test* scores

To investigate possible interactions between these variables, a 3 x 3 x 3 MANOVA was undertaken with sim3, suc3 and exec3 as the independent variables and FRACSCR, WHITESCR and BROWNSCR as dependent variables [Appendix F]. None of the multivariate effects were significant. The effect of sim3 was significant on WHITESCR ($F(2,65) = 3.22$, $p = .046$, $\eta^2 = .061$). Inspection of the means showed that subjects with high ability on simultaneous synthesis tended to rate white music higher than their peers. It seems likely that this result is related to the effect on this variable found for SCHOOL. At school B a group of boys seemed very taken with the white music examples. This possibility is explored below.

Hypothesis 1.7: That on the *Semantic Evaluation of Spectral Density Test*, subjects who attain criterion on the MEK Parts I or V will (a) rate fractal music more positively, and (b) rate white music and brown music less positively, than other subjects.

To test this hypothesis a two-group MANOVA was undertaken with MUSAPT as the independent variable and FRACSCR, WHITESCR and BROWNSCR as dependent variables [Appendix F]. None of the multivariate effects were significant. The univariate effect for FRACSCR was not significant at the 95% confidence level (p = .085). The univariate effects for the other dependent variables were also non-significant. The null hypothesis associated with Hypothesis 1.7 should not be rejected.

The relationships amongst between the variables involved in Hypotheses 1.5, 1.6 and 1.7 were further investigated by computing the correlations between the semantic differential scale scores and the Luria model component scores <u>sim</u>, <u>suc</u> and <u>exec</u> for each of the groups of subjects who had reached, and failed to reach, criterion on the MEK Part I (MEKIHI) and on the MEK Part V (MEKVHI) [Appendix F]. For musically-able subjects, two positive correlations were found between <u>suc</u> and scale scores of ratings of fractal music, and four negative correlations were found between <u>suc</u> and scale scores of ratings of brown music (p < .01) [Table 5.10].

<u>suc</u> by	stale-fresh	unfair-fair	cruel-kind	black-white	ugly-beautiful
fractal music	.41	.40	.73*	.50	.79*
brown music	-.81**	-.70*	-.70*	-.27	-.66*

* p < .01 ** p < .001

<u>Table 5.10 Correlations between successive synthesis and scale ratings of fractal and brown music for subjects who reached MEK criterion</u>

The significant positive correlations suggest that subjects with ability for pitch discrimination employ successive synthesis to encode information for use in favourably rating fractal music on the bipolar scales of 'cruel-kind' and 'ugly-beautiful'. The significant negative correlations suggest that subjects with ability for pitch pattern discrimination employ successive synthesis to unfavourably rate brown music on the bipolar scales of 'stale-fresh', 'unfair-fair', 'cruel-kind' and 'ugly-beautiful'. The bipolar scales of 'cruel-kind' and 'ugly-beautiful' were the most useful for subject ratings, the scale of 'black-white' was the least. In contrast to the earlier MANOVA analysis, these results provide some evidence for the rejection of the null hypotheses associated with the Hypotheses 1.6 (a) and (b). These findings would also suggest that the failure to reject the null Hypothesis 1.7 should be revised, contingent upon individual differences in abilities on successive synthesis.

There were two significant correlations between sim and ratings for white music by subjects who did not attain criterion on the MEK Part 1 [Table 5.11]. These subjects used ability on simultaneous synthesis to rate white music favourably on the scales of 'black-white' and 'ugly-beautiful'.

sim by	stale-fresh	unfair-fair	cruel-kind	black-white	ugly-beautiful
white music	.18	.18	.16	.26*	.28*

* p < .01 ** p < .001

Table 5.11 Correlations between simultaneous synthesis and scale ratings of white music for subjects who failed to reach MEK criterion

This result may reflect the responses of the group of boys referred to above who, by their teacher's report and comments overheard after testing, were enthusiasts for heavy metal rock music. (To this researcher's ear, such rock music is not dissimilar to the randomness which is characteristic of white music.) Such an interpretation was supported by the results of an 2-way ANOVA of sim3 and GENDER on WHITESCR where GENDER was significant (p = .032) [Appendix F]. The rating of white music by boys (mean = 2.0) significantly exceeded the rating of white music by girls (mean = 1.67). On the other hand, this result could also be consistent with the literature which suggests that musically sophisticated subjects employ different cognitive processes in music perception from the musically naive (e.g., Elliot *et al*, 1987; Monahan *et al*, 1987). In a much earlier study, Keston and Pinto (1955) showed that preferences for "serious classical" music were most highly correlated with "intellectual introversion", while preferences for "popular swing" music were most highly correlated with "social extroversion". In any case, this analysis suggests that a semantic differential scale, such as used here, can provide a better account of individual differences in subjective responses to musical stimuli than the one-dimensional scale of "closeness to real music" employed by Voss and Clarke.

A summary of the findings of *Section 5.9* for the research Hypotheses 1.1 to 1.7 is provided in Table 5.12.

Research Hypotheses supported	Research Hypotheses rejected
1.2, 1.3, 1.4, 1.6, 1.7	1.1 (a), 1.1 (b), 1.5

Table 5.12 Summary of Study 1 hypothesis testing

5.10 General discussion

Study 1 investigated the adequacy of information processing within the Luria model, as operationalised by Fitzgerald, to account for individual differences among primary school children on tasks of music perception. Three aspects were investigated, each focussing on different aspects of information processing. First, discrimination of differences in pitch and patterns in pitch and rhythm were measured by the MEK Parts I and V. Second, a musical parallel was sought with the *Fitzgerald Matrices B* marker test for simultaneous synthesis with an original instrument which measured abilities in the recognition of melodic contour inversion. Third, preference for quasi-musical examples generated with three different gradients of the spectral density function - steep, intermediate and flat - was measured on a semantic differential instrument as a more sensitive replication of the study by Voss and Clarke. Overall, in answer to the first research question, the results of this first study provided limited support for the general proposition that individual differences in music perception tasks can be accounted for by abilities on the three information processing dimensions of the Luria model.

More specifically, the third research question asked how individual differences in the perception of pitch patterns might be related to abilities on the Luria model dimensions. The results related to Hypothesis 1.1 showed that, while subjects with criterion level scores on the MEK tests for pitch discrimination and pattern recognition did not have significantly higher mean scores on simultaneous and executive synthesis, abilities on simultaneous synthesis correlated positively with MEK scores when they were considered as a continuous rather than a criterion based variable. The relationship between ability on simultaneous synthesis and MEK scores supports both Karma (1979), that music training increases the capacity for quasi-spatial thinking, and Hassler *et al* (1985, 1987), that music abilities require a high level of simultaneous synthesis to begin with. In contrast, for many of the pre-adolescent subjects in Study 1, cognitive development in the musical domain may still be highly correlated with successive synthesis, and would not have significantly progressed to include a matching contribution by simultaneous synthesis. The influence of musical experience, however, can only be speculated. Certainly, some of the Study 1 subjects would have had a more extensive exposure to music and more experience of music education than their peers, including a greater familiarity with the timbre of the orchestral instruments used in the MEK. It could also be noted that, although neither the MEK Part I nor Part V required identification of absolute or relative pitch, the examples all used diatonic pitch scales, which may have conferred some advantage to those subjects who had more secure scale templates in place. But, Andreasen and Geake (1995) found that some subjects who attained MEK criterion had no prior extra-school music education, and that many with such music experience did not reach criterion.

The second research question asked whether or not a reliable musical test could be developed as an analogue to one of the marker tests already employed in the operationalised Luria model. Hypothesis 1.3 considered the suitability of the *Pitch Contour Inversion Test.* Whereas the results of the item analysis suggested both that simultaneous and executive syntheses contribute to the information processing involved, this task placed a very high demand on successive processing abilities. Dowling (e.g., 1990, 1995a, 1995c) argues that two independent cognitive processes are required for the perception of contour and pitch. It was suggested in Chapter 3 that these cognitive processes could be successive synthesis for contour and simultaneous synthesis for pitch. Young children reproduce the contour of familiar melodies, and only later in their development, if at all, flesh out that contour with accurately pitched intervals. A melodic contour for a person without a pitch template, such as a young child, could be coded as a series of movements in trinary form, up, down or same (Pont, 1990). The relationship between these elements is linear or chained, with an accurate reproduction of the melody requiring their order to be preserved, i.e., utilising successive synthesis.

In the *Pitch Contour Inversion Test* subjects were required to remember up to four phrases for later comparison with a previous phrase. The results suggest that the demands of this task clearly exceeded the capacities for successive synthesis of most subjects, regardless of their musical ability or experience. As Dowling (1971) notes, contour inversion is more difficult to recognise than other change-of-contour tasks such as transposition. Kirby and Das (1990), and Das *et al* (1994), describe a similar "processing bottleneck" with reading [Chapter 2]. Furthermore, the absence of rhythmic emphasis may have increased the difficulty, particularly, perhaps, for those experienced with processing melodies in 'real' music (Nettheim, 1992), or for those whose music processing is more holistic than analytic (Krumhansl, 1992).

This response to the *Pitch Contour Inversion Test* can be interpreted as an outcome of the cyclic hierarchy in the relationships between simultaneous, successive and executive syntheses described in Chapter 2. First, the contour of the target phrase needs to be coded (requiring successive synthesis) and then chunked (simultaneous) for storage in a working memory buffer for later comparison. Information on the actual pitch intervals should be disregarded (executive). This process is repeated during the second hearing of the target phrase, and an initial comparison (executive, successive) might be made to check for fidelity of recall. This sequence of processing events is then repeated four times with each of the different lures, with the additional requirement of inversion of the original contour (simultaneous) to undertake the comparisons between the original contour and both the present and previous lures. Thus, at least five separate successful information processing events might be needed to ensure a successful outcome on each item. As each processing stage presents a filter due to a subject's limitations of ability on the processing dimension required, the likelihood of success becomes increasingly small. The information processing demands of this task are clearly much more stringent than originally conceptualised. The mean score of 6.4 / 15 for subjects with MEK criterion, in the results related to Hypothesis 1.2, is therefore quite commendable.

The results of the *Pitch Contour Inversion Test* are generally consistent with Das *et al* (1994) who point out that processing abilities are domain specific, and Bregman (1994) who argues that perceptual demands are not equivalent across modalities. In particular, the demands of visual perception do not necessarily have parallels with auditory perception. In this task, the successive presentation of the probe and four lure phrases is not surveyable in the manner of the visual pattern inversion task of the computer-adaptive battery where the four distractors are presented simultaneously for comparison. Nevertheless, visualising auditory contour does not seem at the outset to be an unreasonable cognitive demand. For example, it was observed in the pilot study of the *Pitch Contour Inversion Test* that some subjects wanted to draw pitch contours, rather like the probes in the MEK V, to assist them in the task. But, although the MEK Part V required each auditory pattern to be cross-modally transformed for comparison with the visual choices presented on the answer sheet - essentially a task to be undertaken by simultaneous synthesis - there was not a strong relationship between MEK V criterion scores and abilities on this information processing dimension. This suggests that Patterson's (1990) argument for the efficacy of auditory and visual systems for perception of natural complexity needs to be interpreted within and not across modalities so far as individual differences in such perceptual abilities are concerned. These differences in perceptual modalities indicate that the perceptual processing mechanisms at the primary and secondary levels of the second functional block are, as the Luria model suggests, modally specific.

Such an interpretation is also applicable to the results of the Voss replication study where performance on the semantic differential scales would also seem to require simultaneous synthesis, given that the required responses are of a comparative nature. The analyses associated with Hypotheses 1.4, 1.5 (a) and (b), 1.6 (a) and (b), and 1.7 (a) and (b) addressed both the fourth and fifth research questions which were concerned with the preference of children for algorithmically generated fractal music. Here the Luria dimension which was most significant for discrimination between fractal music and brown and white music was successive synthesis, particularly for high musical ability subjects. This is generally consistent with Parker (1978) and others [Chapter 3] that musical aptitude has a strong relationship with other cognitive abilities. As these musics are generated with different degrees of self-coherence, this result is also consistent with Luria's conceptualisation of successive synthesis as being concerned with information as links in a temporal chain (Luria, 1966a, 1973). With non-linear discrete time series such as these fBm examples, Tong (1980) cautions that the degree of complexity is often a function of data length. In this case, data length would be determined perceptually, rather than generatively, by the capacity of a subject's working memory, which, in turn, is directly related to ability on successive processing (Kirby & Das, 1990). In addition, as this task involved discrimination between three separated levels of self-coherence or autocorrelation structure, this finding suggests that some metacognitive facility may be available to high ability music subjects to perceive the degree of linkage of temporal chains composed of various pitches. This interpretation suggests that successive synthesis does operate in the tertiary zones and, in contradiction to Golden (1987), is not necessarily subsumed by simultaneous synthesis at this level. Moreover, such a facility may not be well developed in non-musically able subjects, i.e., superior music ability may involve the employment of different information processing strategies, in accordance with Bever (1988), and in contrast to Bigand (1990) and Madsen *et al* (1993), who argue for uniformity of cognitive processing characteristics for all listeners regardless of musical ability or experience.

An alternate interpretation of the Voss study replication, but not inconsistent with the above, could be made in terms of subject perception of streaming (Bregman, 1994). The three different types of music could be distinguished if streaming was perceivable to a different extent in each type. It could be argued that brown music would present only one stream, whereas white music would present many. It is not clear what might be expected for fractal music; perhaps a couple of streams, or just one stream with a weaker effect. Nevertheless, the preference for fractal music over white and brown replicates the finding of Voss and Clarke with children, and adds weight to the argument that 1/f musics hold the maximum proportion of interest or informational redundancy. However, the absence of stimuli with spectral density coefficients with intermediate values between 1 and 2 leaves the precision of this claim still open. For example, the ratings of fractal and brown music (means = 2.99 and 2.52 respectively) were closer than the ratings of fractal and white music (means = 2.99 and 1.82 respectively). This could be interpreted as being generally consistent with Nettheim's position that the pitch power spectrum of examples of real music are often flatter, i.e., more brown, than the power spectrum of a 1/f pitch series.

Some of the literature reviewed in Chapter 3 indicated that the generation of expectancy and response to novelty is also underpinned by executive synthesis. In this first study, ability on executive synthesis was not correlated with fractal music preference. This is in apparent contrast to Dowling (1990), Jones (1992) and Janata and Petsche (1993) who argue that attentional processes underpin evaluative decisions. However, Bever (1988) argues that music without other features to stimulate a changing attentional focus will not generate evoked emotional responses from listeners. And Boltz (1991) points out that a new style of music will precipitate attentional "down shifting". An outcome of attentional down shifting is an increase in conceptual tempo. This could account for the significant correlations with successive synthesis, which would be more sensitive to an increase in processing speed while listening to music than would simultaneous synthesis.

5.11 Recommendations for a second study

The results from Study 1 indicated at least two directions for a second study. First, the investigation of the relationships between musical ability as determined by the MEK and abilities on the information processing dimensions of the Luria model could be replicated with the use of more sensitive psychometric instruments. The dependence of musical ability on the interaction of the three information processing dimensions might be better understood with the employment of marker tests to measure information processing abilities which are more sensitive than the paper and pencil battery used in this study. The computer adaptive operationalisation of the Luria model (Fitzgerald *et al*, 1995) could fulfil this role. It would also seem that the scores on the MEK tests could provide better measures across a whole sample if used as continuous rather than criterion data, i.e., if MEK scores were used to represent a continuous, rather than a bivariate, variation in music ability.

Second, the replication of the Voss study could be extended to include fractional Brownian motion tone series which lie in between fractal and white music, and in between fractal and brown music. This would require an algorithm to generate tone series with any value of the spectral density exponent $0 < \beta < 2$, i.e., tone series with continuous variation in their autocorrelation from zero to high. In other words, a more sensitive measure of individual differences in music perceptual abilities could be developed around the perception of autocorrelation structure. In Study 1, the superior discrimination shown by musically able subjects between fractal and the other musics was related to successive synthesis. This result is consistent with the suggestion in Chapter 4 that autocorrelation structure is a perceivable characteristic of musical signals which underpins the more readily analytic properties of pitch, timbre, etc., in a cyclic hierarchy involving both simultaneous and successive processing. Certainly such a conjecture requires investigation, which, *inter alia*, might also inform the counter-claims of Gilden *et al* (1993). Thus, two types of perceptual tasks should be designed: one which requires perception of autocorrelation structure mainly by simultaneous synthesis, and a second which requires perception of autocorrelation structure mainly by successive synthesis.

One of the difficulties for interpretation of the results of the first study was the confounding influence of musical experience. It would seem important in a second study to ascertain the musical histories of subjects in terms of music educational experience. To relate music aptitude with general abilities, these data could be complemented with information about school performance in mathematics and English.

Chapter summary

This chapter presented the first study of this research program, an initial limited investigation into the information processing dimensions of musical abilities. Study 1 included a replication of work by Voss on responses to fractal music. The development of instruments and their modification after trialing was described. Multivariate analysis of the data from these instruments, together with data from a Luria model battery and the MEK, was presented. Some evidence was found that the three information processing dimensions of simultaneous, successive and executive syntheses are involved in the perception of musical stimuli such as discrimination of pitch, musical patterns, and pitch contours. The results of Voss were more conclusively replicated, with successive synthesis playing a dominant role in that task. Recommendations for a second study were made.

CHAPTER 6

THE SECOND STUDY

Chapter Overview

Chapter 6 describes the second study which focussed particularly on individual differences in perception of the autocorrelation structure of algorithmic tone sequences. Intact Years 5 and 6 at a government primary school (N = 135) undertook three psychometric batteries:

1. a Luria model battery in computer-based adaptive format (Fitzgerald *et al*, 1995);
2. two original computer-based tests of sensitivity to the perception of autocorrelation structure; and,
3. the MEK Part I *Pitch Discrimination* and Part V *Patterns Recognition* (Bryce, 1979) as administered in Study 1.

Other individual difference data were collected, including music education experience, and performance in school mathematics and language studies. Correlational and multivariate analyses showed that there were significant relationships between the perception of autocorrelation structure and the Luria model dimensions of simultaneous, successive and executive syntheses, as hypothesised. Analyses also revealed significant relationships between success at the perception of autocorrelation structure and success at the MEK tests, and with performance levels at school mathematics and language studies, but no relationship with music education experience. These results are interpreted within the Luria model.

Section 6.1 presents the research questions arising from the recommendations of Study 1. The next section describes two original instruments designed for this study based on a smoothing algorithm of Turner (1992). *Section 6.3* outlines the experimental situation, and *Section 6.4* discusses issues of experimental validity. *Section 6.5* lists the experimental hypotheses derived from the research questions. The sixth section presents the results. A general discussion is presented in *Section 6.7*, while the final section makes recommendations for a third study.

6.1 Formulation of research questions

Chapter 4 presented a detailed analysis of a study by Voss (1975) into the spectral power-law characteristics of music. His investigation had two main conjectures: first, that music follows a 1/f spectral function, like natural fractal fluctuations in time and space; and second, that music generated from 1/f processes was preferred over music from other power spectra. It was further conjectured that human perception may have a preferential acuity for fractal form. The second part of the Voss study was replicated in Study 1 where the results supported the proposition that fractal music is preferred over white and brown musics, particularly by subjects with high pitch discrimination ability, labelled as 'musically able', as measured by the MEK tests.

One of the characteristics by which such musics can be conceptualised to differ is their degree of self-cohesion, or strength of grouping (Bregman, 1994), or level of autocorrelation. It was argued in Chapter 4 that such a conceptualisation was consistent with the degree of information redundancy necessary to maintain musical interest. Moreover, Leman's (1992) model of tone centering proposes that autocorrelation underpins chunking for higher order processing. However, recent work by Gilden *et al* (1993) has called into question whether such musical signal characteristics are indeed perceivable. The first research question is, then: Are tone series with different autocorrelation structures distinguishable from one another?

Given the results from the first study that musically able subjects had stronger preferences for fractal music, and stronger dislikes of the other extremes, the second research question immediately follows: Do musically able subjects have a greater sensitivity to differences in autocorrelation structures?

Evidence from the first study suggested that ability on successive synthesis was important for distinguishing fractal from white and brown musics. As this program of research is interested in the cognitive abilities which underpin musical perception, the third research question parallels the second: Within the Luria model, which processing abilities are important for perceptual sensitivity to differences in autocorrelation structures?

In Study 1, subjects who reached criterion on the MEK tests had higher scores on simultaneous and executive synthesis, but the differences in scores from their peers did not reach statistical significance. This ambivalent result indicates the need for replication. It was anticipated that the new computer adaptive battery would produce a more sensitive measure of abilities on the Luria model dimensions. Also, a better measure of musical abilities across the sample could be achieved by the use of MEK test scores to represent continuous variables as recommended in

Study 1. The fourth research question is, then: Do musically able children have higher measures on any of the three Luria model dimensions?

To assist clarity of interpretation, one of the recommendations from the first study was that both the music educational experience and the general educational achievements of subjects should be taken into account. This raises a fifth research question: Is there a positive relationship between measured music ability and music experience? and a companion sixth research question: Is there a positive relationship between measured music ability and school achievement?

To address these research questions three batteries of instruments were required:

1. a battery to measure musical ability in pitch discrimination, viz., the MEK Parts I and V;
2. a battery to measure abilities on the dimensions of the Luria model, viz., the new computer-adaptive operationalisation by Fitzgerald *et al*; and,
3. a battery to measure perceptual sensitivity to autocorrrelation structure which includes two types of perceptual tasks: one for perception of autocorrelation structure which depends on simultaneous synthesis, and another for perception of autocorrelation structure which depends on successive synthesis. Both these tasks are also likely to depend on executive synthesis.

6.2 Instruments for perception of autocorrelation structure

The main thrust of this study was to investigate whether, and if so the extent to which, the perception of autocorrelation structure in music is an individual difference variable; and if it is an individual difference variable, how much of its variance can be explained by the dimensions of simultaneous, successive and executive synthesis of the Luria model. A limitation on the validity of the *Pitch Contour Inversion Test* of the first study was the problem of memorising musical probe examples, which seemed to create a bottleneck on successive synthesis ability. Consequently, it was important in the development of these new tests that undue demands for successive synthesis were not placed on the subject by requiring him or her to remember more than one musical example at a time.

It was also important that the synthesis of sounds be undertaken by computer algorithm in order to have some control over the way a stimulus varies, and then to be able to relate these changes to variations in reported musical experience. Dowling (1989) offers several suggestions for programming personal computers for experiments in music cognition. A pseudo-code suitable for BASIC programs includes subroutines for randomisation of variables, presentation of trials, and presentation of the experiment. Dowling notes that the pseudo randomisation provided within the resident BASIC is sufficient for an adequate permutation of variables. A similar set of criteria for computer-assisted experiments in music perception is provided by Gibson (1989): an easily adaptable protocol; use of graphics capability; a simple method of subject interaction to minimise data contamination; and, item presentation in an individualised random order. In Gibson's exemplar program, subjects practise on a series of trial items to gain familiarity before the test proper. Item presentation is initiated by the subject, and there is no time limit for subject responses. These suggestions were employed in the instruments described below.

For this second study, an algorithm was required to generate series with a continuous range of values of spectral density. The Landini algorithm was not a suitable candidate for this task, since raising the output values to some index beta in an endeavour to generate other power series causes considerable difficulties with re-scaling the outputs in order to generate music sequences around the same mean pitch. Landini has written an Inverse Fast Fourier Transform algorithm which generates a series for any input value of beta (Landini, 1994). However, because of the extensive calculations involved with each data point, the algorithm is too slow for real time generation of tone series in an instrument for testing sensitivity. This was demonstrated by Spyridis and Roumeliotis (1983) who used FT to 'compose' new melodies, based on a redundancy analysis of forty-five Greek folk songs. Whereas the new compositions showed the same features of the songs that were analysed, having an intermediate value for pitch amplitude redundancy of 37%, their multiple-step analysis was too cumbersome for the real time analysis required here.

Olson and Belar (1961), following Pinkerton (1956), used the frequency distribution of intervals in Stephen Foster songs to drive a digital "music composing machine" which produced new music but which sounded like compositions of Stephen Foster. "Random probability systems allow the generation of sequences of notes in a series which are not completely random nor completely ordered; but each is selected in random fashion with a probability which depends upon preceding notes" (Olson & Belar, 1961: 1163). Noting that entropy is a measure of the randomness in a system, these authors define melody as "an agreeable succession or arrangement of sounds" (p. 1164), which implies certain patterns of regularity. They consequently argue that

> *the entropy of the melody must be sufficiently low so that a definite pattern is established. On the other hand, the entropy must be high enough to incorporate sufficient complexity to provide a degree of sophistication. From the foregoing, the general conclusion may be drawn that the musical tones of a melody possess a relatively low value of entropy* (p. 1164).

Turner (1992) generalised Olson and Belar's approach to generate stochastic "melodies" from an algorithm for exponential smoothing which produces a numeric series having a power spectrum of $1/f^{\beta}$

$$Y_t = Y_{t-1} + A\,(R - Y_{t-1}) \qquad (6.1)$$

where Y_t = series value at "time" t, R = random number, and $0 < A < 1$ = smoothing coefficient. For any particular series, β is not a strict function of A; rather the relationship is statistical. Turner reports means for 100 series which show an inverse linear relationship between β and A [Table 6.01]. For example, $A = 0.05$ gives mean $\beta = 1.62$, while $A = 0.90$ gives mean $\beta = 0.15$. In the limits, $A = 0$ produces an unchanging series with maximum autocorrelation. $A = 1$ produces a purely random series (white music, $\beta = 0$) with zero autocorrelation. The suggested value for music, $\beta = 1$, is approximated by the Turner algorithm with series where $0.4 < A < 0.5$.

An advantage of the Turner algorithm over an IFT approach is that the smoothing coefficient A is a continuous and monotonically increasing variable over the range. That is, it offers an opportunity to quantify Olson and Belar's "relatively low value". However, decreasing values of Pearson r in Table 6.01 could indicate that the validity of A as a measure of β is lower for the less correlated series. This is supported by Landini (1994) who presents a plot of β vs A of 990 sequences each using 1024 data points [Figure 6.01]. Inspection of this plot suggests a dominant linear relationship for $0.2 < A < 0.9$, i.e., $0.1 < \beta < 1.3$, and linearity with a slightly higher gradient for low values of A, $1.3 < \beta < 2$.

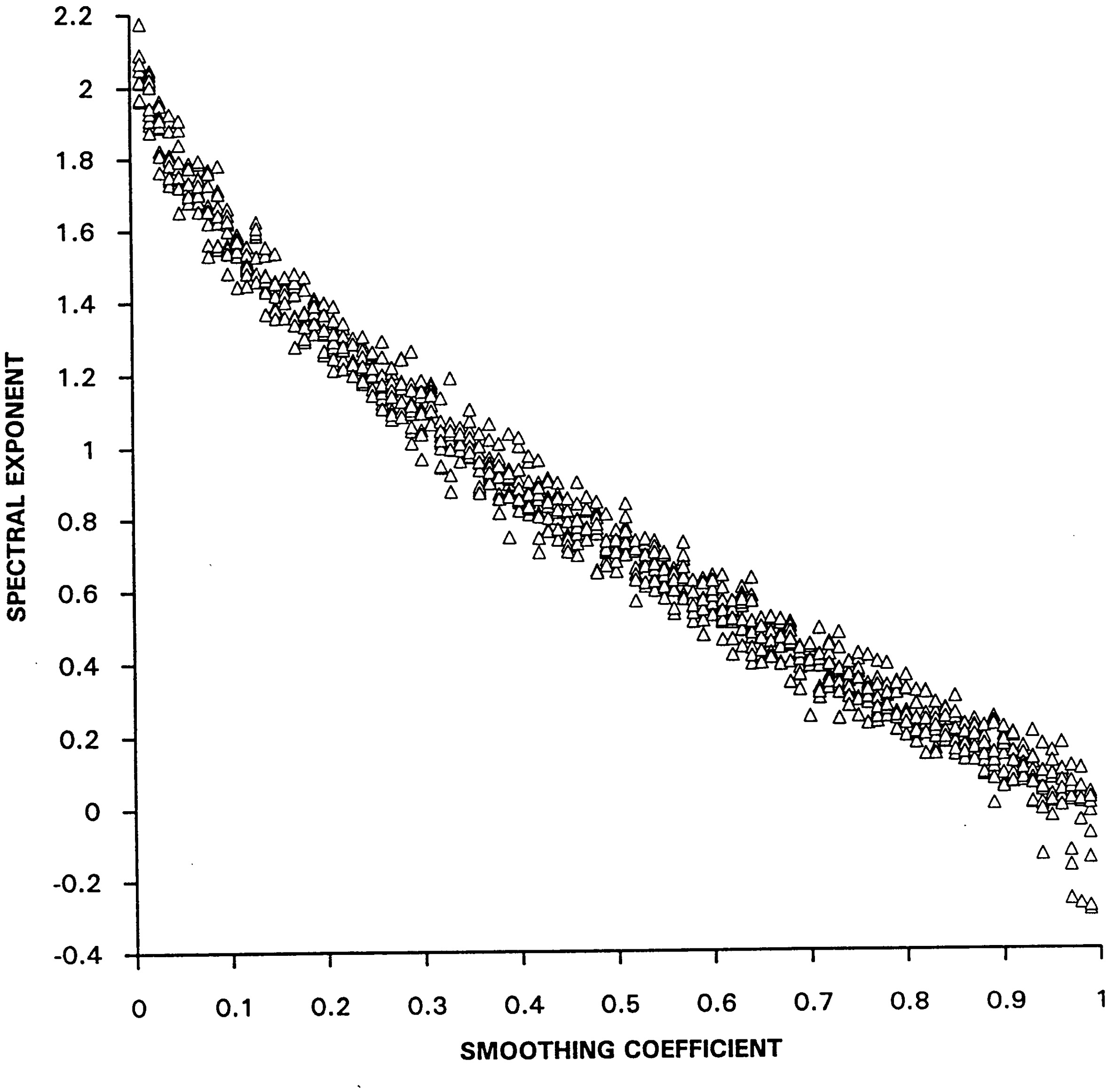

Figure 6.01 Spectral power exponent vs. smoothing coefficient plot
(Landini, 1994)

Smoothing coefficient A	Spectral power exponent β	Pearson correlation r
.05	1.62	.98
.10	1.64	.98
.20	1.40	.97
.30	1.22	.97
.40	1.04	.96
.50	.90	.94
.60	.70	.90
.70	.54	.86
.80	.31	.75
.90	.15	.67

Table 6.01 Relationship between smoothing coefficient and spectral power function exponent (Turner, 1992: 19)

As these departures from linearity were not severe, A was taken as an indicator or approximate measure of the strength of autocorrelation of the series generated by the Turner algorithm. The strength of this relationship between A and the autocorrelation structure was assessed by computing the autocorrelation functions of outputs of the Turner algorithm for $0.05 < A < 0.95$, $\Delta A = 0.05$ [Table 6.02]. Inspection of Table 6.02 shows that the log gradients of the autocorrelation function to the first zero crossing are monotonically increasing for increasing A. The number of lags before the first zero crossing, however, are not monotonically decreasing. Together these suggest some limitations on the robustness of this algorithm, but not of a severity to override its pragmatic value to the present study.

To generate "melodies" each Y_t in a series determined the number of semitones in the corresponding interval above A2 = 110 Hz. The random number distribution was set so that the highest note was about A5 = 880 Hz. The semitone ratio = 1.05943 for equal temperament was used (Johnston, 1989). It produced frequencies which agreed with music tunings to two significant figures. Each series was of 50 notes duration at a constant tempo of 10 Hz and constant volume. The musical scores of three examples, A = 1, A = 4.5 and A = 9, are shown in Figure 6.02. Graphically these scores are similar to short sections of the fBm series with different Hurst exponents [Figure 4.02] where A is inversely related to H. The pitch distributions of the series for A = 1 and 4.5 sound similar to the pitch distributions of the algorithmic brown and fractal musics [Chapter 4]; the A = 9 series sounds somewhat more unpredictable than algorithmic white music. The sequences are longer than the two to six tones that subjects require to detect similarities or differences (Karma, 1985). The tempo is in the range suggested by Pennycook (1985) for the streaming of performance data in computer-music.

Smoothing coefficient A	log gradient autocorrelation	lags before first zero crossing
.05	.142	10
.10	.196	9
.15	.226	8
.20	.249	8
.25	.273	8
.30	.299	8
.35	.329	8
.40	.359	8
.45	.391	8
.50	.421	6
.55	.449	6
.60	.475	6
.65	.498	6
.70	.518	6
.75	.535	6
.80	.549	6
.85	.561	6
.90	2.30 *	2
.95	3.30 *	2

(* gradient taken from n = 0 to 4 except for A = .90, .95.)

Table 6.02 Relationship between smoothing coefficient and the autocorrelation function of Turner algorithm outputs

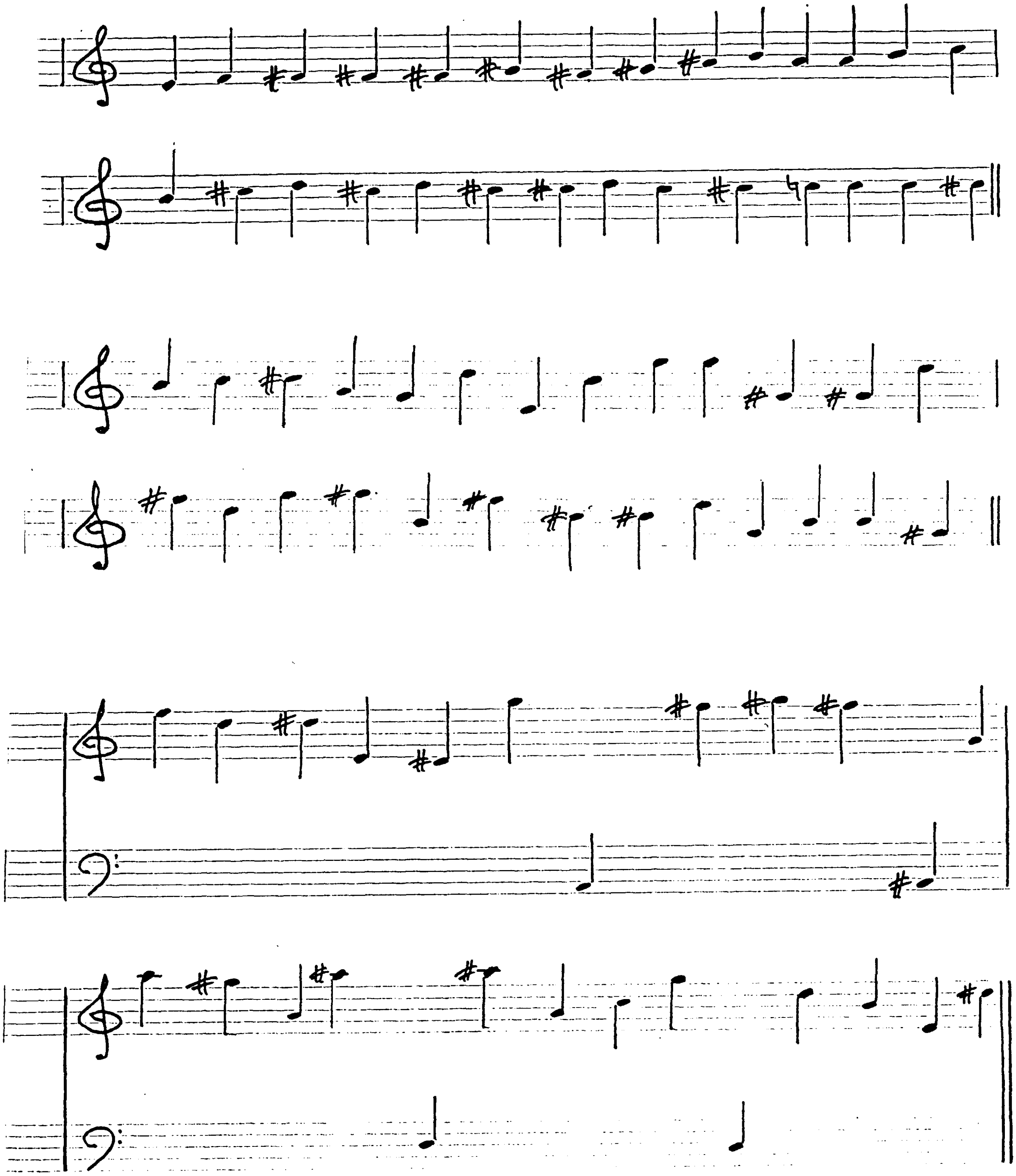

Figure 6.02 Tone series for three different smoothing coefficients

top: A = 1 middle: A = 4.5 bottom: A = 9

It should be noted that there is no claim for necessary musicality of the output tone series. The construction of intervals above a bottom note is consistent with the score reduction technique of Hsu (1993), which, it was argued in Chapter 4, ignored keyality and harmonic modulation. Both Roads (1985) and Pennycook (1985) point out the limitations of statistical techniques in computer generated music. Pennycook argues that the "rich, descriptive vocabulary of music concepts" (p. 268) prohibits the ready translation of music information into computer data. Roads portrays a future where computer programs could not just "find instances of musical structures", but make analytic connections between such structures in the pursuit of understanding. A prerequisite would be "the ability to infer concepts from partial descriptions", which is a focus of this study.

In contrast to Balch (1989), for example, the Turner tone series are atonal. Balch argues that diatonic tonality is a better basis for test items because of its familiarity. A counter argument suggests that familiarity with diatonic tonality is a function of music experience, and thus atonality may reduce the variance due to such experience by placing more and less musically experienced subjects on equally unfamiliar ground (McMullen, 1974), although Krumhansl, Sandell and Sergeant (1987), Crummer, Hantz, Chuang, Walton and Frisnia (1988) and Tsuzaki (1991) found that experienced musicians had superior discrimination on atonal stimuli, possibly due to greater reliability of pitch memory. On the other hand, Miyazaki (1993) argues that reliable pitch memory can be dysfunctional in some non-tonal contexts.

A more potentially serious compromise to musicality is the absence of rhythmic variation and emphasis. However, Boroda (1993), Nettheim (1992), and Schmuckler and Gilden (1993) have shown that rhythmic variation, while a vital dimension of melody, does not usually follow the same inverse power law function as pitch and melody. There would seem to be a gain in experimental control, then, by employing uni-rhythmic stimuli.

This adapted Turner algorithm was used to generate tone series for two experimental tasks. The first, *Scale Estimation of Smoothing Test* [Appendix G], was an auditory parallel of an instrument of Pentland (1984) who measured perception of roughness as a function of fractal dimension with computer-generated landscapes of different values of D [Chapter 4]. Radocy and Boyle (1988) operationalised music perception as a scale fitting task - matching a sensory continuum to a number scale. The similar results from both children and adults suggested that the measured effects were not learned. "Individuals will vary in their sensations and in their abilities to express them with numbers" (p. 61). The *Scale Estimation of Smoothing Test* required subjects to estimate the smoothing coefficient A normalised on a scale from 0 to 10, after listening to a tone series once. Estimates could be given to one decimal place. Feedback was provided on each of 20 items, with each A selected by a random number algorithm. Before the test was undertaken, several examples were played with their A values displayed. Next, subjects were asked to label each end of the scale after listening to examples with A = 1 and A = 9. The labels were recorded as a qualitative measure of perceptual sensitivity. After labelling, subjects could practise scale estimation with feedback until they were ready to take the test. The instrument was trialled on twelve Year 6 subjects from schools other than the school which participated in the study, and included subjects with high music skills. All subjects in the trial reported perceived differences between A = 1 and A = 9 series, and estimates for most subjects generally followed the variations in A, although in two cases subjects estimated 10 for a monotone where A = 0. On the other hand, in a couple of cases subjects estimated A correctly to one decimal place.

Analysis of various summary statistics suggested that the root-mean-square difference,

$$\mathrm{RMSD} = \Sigma\ \frac{\sqrt{(A_i - E_i)^2}}{n} \qquad (6.2)$$

where A_i = smoothing coefficient of item i, E_i = subject estimation of A_i, and n = number of items, gave the best measure of individual performance. This statistic was computed for all subjects in the study. The value of RMSD at chance or guessing, i.e., for random E_i, was calculated: RMSD chance $\approx$ 4.04. This is assuming that the estimation task is linear, i.e., that scale estimation is equally easy or difficult at all points on the scale. The validity of this assumption was checked by plotting RMSD vs A for each item for all trial subjects. The data plot approximated a filled circle indicating no relationship between degree of difficulty and position on the scale, unlike the task of Schmuckler and Gilden (1993) [Chapter 4]. The linearity assumption was further checked with the results of the responses to the second instrument detailed below.

Because a scale is a quasi-spatial construct, it was conjectured that performance on this task would require a high level of simultaneous synthesis. In the instrument trial it was noted that RMSD was not related to music education experience but was negatively correlated with school performance. That is, task performance was positively related to academic performance.

The second task, *Change of Smoothing Test* [Appendix H], adapted a phrasing task of Serafine (1988) where subjects had to nominate the boundary point between two successive musical phrases. Narmour (1990) suggests that experiments for expectation of continuation could investigate the separate processes of implication and realisation. Schmuckler (1990) argues that, because melodic processes occur across all time scales, higher levels of coherence generate clearer expectancies for future events, especially endings. Balch (1981) notes that the redundancy of two part-melodies was an important factor in determining subject ratings of how well the second part followed the first. The *Change of Smoothing Test* required subjects to indicate when they thought a series changed its smoothing coefficient from one value of A to another. A central measure in the *Change of Smoothing Test* was the response latency in note count from the moment of change to the moment when the subject indicated perception of the change by striking any key on the computer keyboard. A net latency was calculated by subtracting the response latency to an obvious change from 220 Hz to 880 Hz, averaged over five trials. The response latency was typically four or five counts. This initial response exercise was used as an introduction to the perception of autocorrelation battery. Each test item consisted of two concatenated adapted Turner algorithms with different values of A. Although the values of A were randomly determined, the differences were constrained to begin at a minimum of 6 out of a maximum of 10, and progressively decrease as the test proceeded. This, it was conjectured, should make the test progressively more difficult. The length of each item was 100 notes, but the length of the first series was randomised between 35 and 75 notes. The length of the second series made up the difference to 100.

The test begins with several examples where the commencement of the second series is prominently indicated by changes to both the displayed graphic ("1" to "2") and the screen colour (blue to red). Next, examples are played for subject listening with 'obvious' changes: A = 9 concatenated by A = 1, and A = 2 concatenated by A = 8. Subjects can then practise, with feedback, until they are ready to take the test. Feedback is given on each item of the test, either as the net latency, or as indications that the subject responded too early, or did not respond at all. As a baseline control for this last response condition, three items did not change. Here a non-response received feedback that this was, in fact, a correct response.

Balch (1981) notes that response to continuation tasks is dependent upon cognitive processing, and in particular, is determined "by the subject's ability to retain both parts long enough to detect the presence of a relationship between these parts" (p. 49). Thus it could be expected that performance on the *Change of Smoothing Test*, requiring the interruption of a data sequence, would be dependent on ability on successive synthesis. However, some degree of pre-attentive processing may be necessary to establish a sense of coherence of the data sequence. The thrust of Serafine's (1988) studies was to address the degree to which "children's scanning or monitoring of an ongoing phrase ... is general and global or focussed on the particularities of local, low-level events" (p. 112).

The test was trialled during development with the same subjects as the *Scale Estimation of Smoothing Test*. Trial subject responses determined the final range of difficulty, the limits on series length, and the number of items which was set at 33 in order to measure a reliable degree of sensitivity. Inspection of the net latency responses of the trial subjects showed a bimodal distribution. Latencies were either small, indicating a correct response, or larger than 25 indicating a guessed response that a change should have occurred by then. A number of candidate statistics were subsequently generated: mean latency (XLAT), number of correct responses (NGOOD), number of late responses or guesses (NGUESS), number of early responses (NEARLY), and, number of correct responses to controls (NCONT). The response to each of the controls was noted (CONT1, CONT2, CONT3), as well as the value of A (ACONT1, ACONT2, ACONT3). Since the trial results of the *Scale Estimation of Smoothing Test* indicated that degree of perceptual difficulty was not related to the size of smoothing coefficient A, it was assumed here that the degree of difficulty of perceiving changes in A would be related simply to the differences in A for the two tone series. Consequently, the weighted latency (XWTLAT) was calculated,

$$\mathrm{XWTLAT} = \mathrm{XLAT}\,|A_1 - A_2|. \tag{6.3}$$

For the trial subjects XLAT seemed to provide the best measure of individual differences on performance. However, all statistics were recorded for the second study.

6.3 Experimental situation

The subjects (N = 135) were 10-11 year olds from Years 5 and 6 from the North Coast government primary school which was used in Study 1 (School B). The school was selected for internal diversity of socio-economic profile in order to maximise generalisability of the results, and for the positive support of teachers and the principal. The four teachers of the participating classes and the school principal were briefed on the aims and methodologies of the research. A permission note explaining the research was sent home to the families of all subjects. No children were withdrawn by parental non-permission. None of the subjects participated in the first study. The age band (10-12 years) of the subjects was chosen to match that used in Study 1. Testing was undertaken over a six week period to allow time for all subjects to individually undertake both the computer-based sensitivity to autocorrelation structure and Luria model batteries. The MEK Parts I and V were administered in the third week of testing. There was some subject mortality of children who left the school during the testing period, resulting in N = 126 for the Luria model tests, and N = 132 for the autocorrelation tests. The computers used were located in classrooms and the school library, and were very familiar to all subjects.

Each subject was asked about his/her music experience with a questionnaire that used at least one question from each of McPherson's (1993) four factor scales of 'early exposure', 'enriching activities', 'length of study' and 'quality of study'. The questions were:

1. Do you play an instrument or sing?
2. Which instrument(s) do you play?
3. Do (did) you have music lessons outside of school?
4. For how many years have you been playing your instrument?
5. Do you sit for Australian Music Examination Board (AMEB) exams?
6. What is the highest Grade you have achieved?
7. What level (Pass, Credit, Distinction) did you pass this Grade at?

Responses were used to generate a six-point *Music Experience Scale* (labelled MUSICEXP). The scale provides a cumulative rating of music experience, i.e., higher ratings indicate a level of experience and achievement which subsumes the levels indicated by lower ratings. A rating of "0" was recorded for no reported music experience beyond that provided by the school. The Years 5 - 6 Creative and Performing Arts (CAPA) Syllabus has a mandatory music component which the school pursues with enthusiasm through the employment of part-time music specialist teachers. The program includes choral singing, and some rudimentary keyboard and recorder playing. Those subjects who indicated some past extra-school music activities which were not maintained were generally rated "1"; if past music experiences included formal music lessons then the rating was "2". Subjects who reported playing at home or with friends beyond school were also rated "2". Those subjects who had private music lessons for several years were rated "3". Subjects were rated "4" if they had achieved an AMEB grade commensurate with the usual grade for age 10-12 years, i.e., Grade 2 or 3 (McPherson, 1993). The highest rating of "5" was reserved for those few subjects who played two or more instruments, one at least at AMEB Grade 4 or higher. 'Borderline' ratings were promoted if the instrument was keyboard or an orchestral string instrument.

The four class teachers were asked to confidentially rate each of their subjects on a normative scale for achievement in school mathematics (labelled MATH) and English (labelled LANG). This four-point scale was used some years ago by secondary schools for streamed placements of Year 7 entrants. As such it is a scale with which experienced teachers are familiar, and ratings could be expected to be reliable. Unfortunately for the purposes of this research, the distribution of grades, "A" (best) about 40%, "B" about 30%, "C" about 20%, and "D" about 10%, better discriminates at the lower end of the scale. Also, two of the teachers began their teaching after the use of this scale was discontinued. They were asked to provide performance estimates on a ten-point scale. These estimates were then collapsed on to the four-point scale. The teachers subsequently confirmed the validity of these estimates by comparison with the results of Basic Skills Tests in Mathematics and English administered by the NSW Board of Studies.

6.4 Experimental validity

The experimental situation was similar to the first study in terms of internal and external validity. N = 135 is sufficiently large to minimise Type II error in designs with small effect size and ensure an adequate minimum cell size (Stevens, 1986). To undertake this research in a school, intact classes had to be used. As tests were administered to the whole Year, and all subjects were administered the same battery, diffusion between experimental groups was not an issue. The MEK battery was administered by the same person to both Years as a whole. The sensitivity and Luria model batteries were computer-based, so there was no possible effect for tester. The *Scale Estimation of Smoothing Test* and *Change of Smoothing Test* were undertaken with headphones so as not to interrupt the class, and to minimise confounding aural distractions. There were no severely hearing impaired children among the subjects. The adequacy of the language used for the test items and the instructions was confirmed by the children used as subjects for the trialing of the instruments, and the classroom teachers of subjects participating in the study. There were no cases of tests being incomplete due to misunderstanding of instructions. A research assistant read the instructions aloud to subjects with low reading skills.

The six week test period enabled all subjects to be tested, and thus controlled some of the many extraneous variables which arise from testing in schools, such as organised absences through school camps during the testing sessions, special events such as school concerts, and unforeseen absences due to contagious illnesses. The selection of the school was determined by the supportive attitudes of the teachers towards research, and the parents' interest. Testing did not increase the school work load, as it was undertaken during normal lesson times. The subjects commented on the novelty of all the test instruments used. Many subjects commented on their enjoyment of the tests, particularly the 'sensitivity to autocorrrelation structure' battery. From teachers' comments on the novelty of computer-based testing, there should be no effect of familiarisation with this style from school testing unassociated with this research.

As with Study 1, any extraneous variables, such as they were, would reduce the effect size rather than produce a spurious result. Thus it was the adequacy of the testing outcomes, more than internal validity, that was at issue.

6.5 Experimental hypotheses

The six Research Questions in *Section 6.1* were formulated into six Experimental Hypotheses.

Hypothesis 2.1: That there are positive relationships between success on the *Scale Estimation of Smoothing Test* and measures of (a) simultaneous synthesis, and (b) executive synthesis.

Hypothesis 2.2: That there are positive relationships between success on the *Change of Smoothing Test* and measures of (a) successive synthesis, and (b) executive synthesis.

Hypothesis 2.3: That there will be positive relationships between scores on the MEK and (a) success on the *Scale Estimation of Smoothing Test*, and (b) success on the *Change of Smoothing Test.*

Hypothesis 2.4: That there will be positive relationships between scores on the MEK and measures of (a) simultaneous synthesis, (b) successive synthesis, and (c) executive synthesis.

Hypothesis 2.5: That subjects with high degrees of success on the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* will have more music education experience than subjects with less success.

Hypothesis 2.6: That subjects with high degrees of success on the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* will have higher school achievement levels in mathematics and English than subjects with less success.

6.6 Data analysis

All data were analysed with *SPSS for Windows* [Appendix I]. Significances are reported at the 95% confidence level unless otherwise indicated. ANOVA procedures were used to investigate effects on several criterion variables individually in Hypotheses 2.1 and 2.5; MANOVA procedures were used to investigate effects on a number of criterion variables as a set in Hypotheses 2.2, 2.4, 2.5 and 2.6.

Preliminary analysis

The set of six Luria model marker tests were reduced to three orthogonal dimensions using component analysis with Varimax rotation [Table 6.03]. The loadings of the two marker tests for successive synthesis, *Number Span* (NUMBER) and *Word Span* (WORD) on Component 1 indicated that this component reflected successive synthesis. The loadings of the two marker

	Component 1	Component 2	Component 3
INVERT	.07	**.77**	.38
PAPER	.26	**.87**	.06
NUMBER	**.87**	.14	.12
WORD	**.78**	.15	.26
SIZE	.22	.21	**.88**
NUMLET	**.58**	.25	**.56**

Table 6.03 Principal components structure of Luria model marker tests for Study 2

tests for simultaneous synthesis, *Inverted Matrices* (INVERT) and *Paper Folding Test* (PAPER) on Component 2, indicated that this component reflected simultaneous synthesis. Similarly, Component 3, with loadings of *Size Attention Span* (SIZE) and *Letter/ Number Attention Span* (NUMLET) reflected executive synthesis. The similar loading of NUMLET on to Component 1 was indicative of the successive demands of this task and its similarity to the two marker tests for successive synthesis. Bartlett component scores (sim, suc and exec) were calculated for further analysis and also used to partition subjects into high (sim1, suc1, exec1), medium (sim2, suc2, exec2) and low (sim3, suc3, exec3) scores on each of the three components.

The variables from the *Scale Estimation of Smoothing Test* and *Change of Smoothing Test* were tested for normality of distribution by computing the coefficients of kurtosis and skewness [Table 6.04]. For RMSD there was an acceptable degree of platykurtosis and an acceptable degree of positive skewness; the distribution of this variable was a good approximation to a normal distribution. For the distributions of the *Change of Smoothing Test* variables, the coefficients of skewness and kurtosis indicated that the distributions of NEARLY and NCONT were somewhat positively skewed.

To examine the possibility that school Year 5 or 6 (FORM) related to the variables considered in this analysis, a two-group MANOVA for FORM as the independent variable was undertaken on the Luria model component scores sim, suc and exec, the criterion variables to the two perception of autocorrelation structure tasks, RMSD, XLAT, XWTLAT, NGOOD, NGUESS, NEARLY

sensitivity to autocorrelation structure	test variable	coefficient of skewness	coefficient of kurtosis
Scale Estimation of Smoothing	RMSD	.437	2.8
Change of Smoothing	XLAT	-.66	3.62
	XWTLAT	.22	4.26
	NGOOD	-.59	2.07
	NEARLY	1.09	4.41
	NGUESS	.48	2.90
	NCONT	1.02	3.23

Table 6.04 Coefficients of skewness and kurtosis of sensitivity to autocorrelation structure variables

and NCONT, the scores on the MEK Parts I and V, MEK1 and MEK5, and the three estimates of educational performance, MUSICEXP, MATH and LANG, as dependent variables [Appendix I]. The multivariate effect was significant (Wilk's $\lambda = .642$, $p < .001$, $\eta^2 = .358$). The significant univariate effects are reported in Table 6.05. These results may be due to effects from testing since each Year was tested separately, or, they may indicate that performance on the perception of autocorrelation structure and MEK Part I tasks was partly determined by age or experience.

DV	F	df	significance p	effect size η^2
RMSD	4.67	1,105	.033	.043
XLAT	5.72	1,105	.019	.052
NCONT	5.29	1,105	.023	.048
MEK1	20.77	1,105	< .001	.165

Table 6.05 Significant univariate effects of School Year for Study 2 variables

To test this interpretation of the MANOVA, Pearson correlations between the continuous variable AGE and all of the criterion test variables were also computed. None were significant except for NCONT and MEK1 [Table 6.06]. The significance for NCONT was probably a spurious effect, although it might suggest that younger subjects were less confident about deciding when a change occurred, and so did not respond at all to several items including the controls. The significant correlation between subject age and the MEK Part I, a test for pitch discrimination, was consistent with Billingsley and Rotenberg (1982) who found that interval discrimination was related to age up to 12 or 13 years.

Variable	Age
SIM	.02
SUC	-.07
EXEC	-.07
RMSD	-.12
NGOOD	-.16
NGUESS	-.09
NMISS	-.04
NEARLY	.10
NCONT	-.19*
XLAT	-.16
XWTLAT	-.10
MEK1	.34**
MEK5	.04
MUSICEXP	.01

* p < .05 ** p < .001

Table 6.06 Correlations between age of subjects and component scores and test variables

Analysis of Contrasts

To further investigate the role of simultaneous, successive and executive syntheses in the perception of autocorrelation structure, pitch and pattern perception, and school performance, a 3 x 3 x 3 MANOVA with planned contrasts for linear (contrast 1) and curvilinear effects (contrast 2) of sim3, suc3 and exec3 as the independent variables was undertaken with all Study 2 criterion variables RMSD, the set of NGOOD, NGUESS, NEARLY and NCONT, MEK1 and MEK5, and MUSICEXP, MATH and LANG, as dependent variables [Appendix I]. A MANOVA design with these planned contrasts separates the effects due to linear variance from the effects due to curvilinear variance. The coefficients are set out as follows:

	high	middle	low
extremes	1	0	-1
curvilinear	-1	2	-1

The more rigorous confidence level of 99% was adopted to reduce spurious chance effects arising from this design.

The significant multivariate effects are reported in Table 6.07. The difference between high and low abilities on successive and executive syntheses together accounted for over 60% of the variance in performance on this battery of tasks, indicating the importance of these two dimensions of information processing in particular for tasks of music perception. None of the curvilinear contrasts were significant at the multivariate level. The univariate effects are reported with the results of the related hypothesis testing.

multivariate effect	Wilk's λ	significance, p	multivariate effect size
suc3 (1) x exec3 (1) (linear)	.738	.011	.262
suc3 (1) (linear)	.676	.001	.324
exec3 (1) (linear)	.688	.001	.312

Table 6.07 Significant multivariate effects for linear contrasts of 3-way partitions of Luria model component scores on Study 2 variables

Hypothesis 2.1: That there are positive relationships between success on the *Scale Estimation of Smoothing Test* and measures of (a) simultaneous synthesis, and (b) executive synthesis.

The variable RMSD, the measure of performance on the *Scale Estimation of Smoothing Test,* had a mean = 3.06. The value for a random response = 4.04. A one-sample t-test showed that the difference was significant ($p < .0001$, one-tailed), indicating that average performance was better than that achieved by guessing. In other words, the results of the *Scale Estimation of Smoothing Test* provided non-random information.

To test the Hypothesis 2.1 (a) and (b), two-tailed correlations were computed between the Luria model component scores sim, suc and exec, and RMSD [Appendix I]. The correlation was significant for sim ($r = -.34$, $p < .001$) [Table 6.08]. It should be recalled that RMSD is an error type term where low scores indicate superior performance. Thus the negative value of r indicated that success on the *Scale Estimation of Smoothing Test* was positively correlated with ability on simultaneous synthesis. The correlation between exec and RMSD approached significance [Table 6.08]. Again, because RMSD is an error term, the negative value of r indicated that success on the *Scale Estimation of Smoothing Test* was positively correlated with ability on executive synthesis. Measures of Luria model component scores together accounted for 15% of the variance (Σr^2) of RMSD. The null hypothesis associated with Hypothesis 2.1 (a) should be rejected. The evidence is not strong enough to reject the null hypothesis associated with Hypothesis 2.1 (b).

RMSD by	Pearson r	significance, p	effect size, r^2
sim	-.34	<.001	.116
suc	-.22	.016	.048
exec	-.16	.077	.026

Table 6.08 Correlations between Luria model component scores and *Scale Estimation of Smoothing Test* scores

Further support for these results was found from the univariate tests associated with the planned contrasts MANOVA [Appendix I, Table 6.09]. The linear effect of sim3 (1) was significant for RMSD, supporting the hypothesis that performance on the *Scale Estimation of Smoothing Test* is dependent on simultaneous synthesis. The effect for the curvilinear contrast of sim3 (2) was not significant. The linear effect of suc3 (1) approached significance for RMSD, suggesting that successive synthesis might play an unhypothesised role in performance on the *Scale Estimation* *of Smoothing Test*. The effect for the curvilinear contrast of suc3 (2) was not significant. Neither of the effects for the contrasts exec3 (1) and exec3 (2) were significant.

univariate effect	F	df	significance p	effect size η^2
sim3 (1) linear	8.26	1,97	.005	.078
suc3 (1) linear	6.10	1,97	.015	.059
exec3 (1) linear	3.63	1,97	.060	.036

Table 6.09 Significant univariate effects of Luria model component scores and *Scale Estimation of Smoothing Test* scores

The means of RMSD for the high-medium-low partitions of sim3 and suc3 are reported in Table 6.10. Noting that RMSD is an error term, the monotonic increase in the means across the high-medium-low partitions indicated that subjects with high ability on simultaneous synthesis, and subjects with high ability on successive synthesis, were more successful at the task of estimating the smoothing coefficient.

RMSD	partition	sim3	suc3
	high	2.68*	2.84
	medium	3.08	2.92
	low	3.40*	3.40

* significant linear effect

Table 6.10 Means of *Scale Estimation of Smoothing Test* scores for high-medium-low partitions of simultaneous and successive synthesis

These results indicate that the task of making scaled estimations of autocorrelation structure requires both successive and simultaneous syntheses, with around 14% of the task variance explained by differences between high and low information processing abilities. This result is consistent with the Luria model where information received and initially processed by successive synthesis is then chunked for further processing by simultaneous synthesis.

In the *Scale Estimation of Smoothing Test* subjects were asked to label the extremes of the scale after listening to two examples with the smoothing coefficient A = 1 and A = 9. Examples of these labels are given in Table 6.11, along with RMSD scores, and sim3, suc3 and exec3 partitions. The responses in Table 6.11 are indicative of the variety of performance on the *Scale Estimation of Smoothing Test.* It was conjectured that subjects who employed useful labels would perform better on the scale estimation task (example cases 1, 2, 5, 9; counter example cases 6, 10, 12, 13).

case	A = 1 label	A = 9 label	A ZERO	RMSD	sim3	suc3	exec3
1	close	far	0	1.20	1	1	3
2	kept-in	jumpy	0	2.87	1	3	1
3	flat	jumpy	0	4.80	3	2	1
4	boring	wild	0	3.85	3	3	3
5	together	untogether	0	2.04	1	1	1
6	happy	jazzy	5	3.36	3	2	2
7	flat	wild	6	3.01	1	1	1
8	flat	interesting	8	2.64	2	1	3
9	smooth	bumpy	9	3.85	1	3	2
10	dum	wierd	10	3.74	3	1	3
11	fast	pop	10	2.93	3	1	2
12	fast	reapetin (sic)	10	3.80	3	3	1
13	up and down	stupid	10	5.10	3	3	3
14	harmonious	technotronic (sic)	.	1.70	2	1	1
15	busy	robotic	.	3.31	1	2	2
16	fast	active	0	3.08	3	1	2

Table 6.11 Selected responses to *Scale Estimation of Smoothing Test*

In many cases, one of the items had A = 0, i.e., an unchanging tone. Response to this item (labelled AZERO) is also recorded in Table 6.11. It was conjectured that response to this particular item might also be indicative of overall task performance (example cases 1, 2, 5; counter example cases 10, 11, 12, 13). A two-tailed correlation between responses to the A = 0 item and RMSD was significant ($r = .45$, $p < .001$). To some extent this could be expected since the penalty for estimating A = 0 at the other end of the scale, say, at 9 or 10, adds about 0.5 to the value of RMSD. However, inspection of the data showed that only 10.5% of subjects went to this extreme. Some 58.7% of subjects estimated this item at 1.0 or less, including 40.5% who correctly estimated 0.

Table 6.11 also indicated how simultaneous synthesis was important to consistent performance on this task over the twenty items (example cases 1, 2, 8, 14; counter example cases 3, 4, 13). Subjects low on simultaneous synthesis, nonetheless, could attain success by presumably utilising strategies involving their higher abilities on successive synthesis (cases 8, 11, 16) or executive synthesis (cases 2, 16).

In Study 1, the *Pitch Contour Inversion Test* did not appear useful as a marker test for simultaneous synthesis in the music domain. Here, the possibility that the *Scale Estimation of Smoothing Test* might fulfil such a role was explored. A principal components analysis with Varimax rotation was undertaken with the component scores sim, suc and exec, and the *Scale Estimation of Smoothing Test* variable RMSD [Table 6.12]. A three component structure was

	Component 1	Component 2	Component 3
SIM	**.89**	-.15	-.11
SUC	.02	**.96**	-.05
EXEC	.02	-.04	**.98**
RMSD	**-.71**	**-.38**	-.27

Table 6.12 Component structure with *Scale Estimation of Smoothing Test* scores

recovered. Component 1 with a high loading of sim also had a high loading of RMSD. This result suggests that the *Scale Estimation of Smoothing Test* could possibly act as a marker test for simultaneous synthesis which, unlike other such marker tests, utilises auditory stimuli. However, its non-negligible loading on Component 2 with suc compromises its utility in this regard.

Hypothesis 2.2: That there are positive relationships between success on the *Change of Smoothing Test* and measures of (a) successive synthesis, and (b) executive synthesis.

To test Hypothesis 2.2 (a), two-tailed correlations were computed between the Luria model component scores suc and exec, and each of the variables XLAT, XWTLAT, NGOOD, NGUESS, NEARLY and NCONT as measures of performance on the *Change of Smoothing Test* [Appendix I]. The correlations between suc, exec and NGOOD, and between exec and NEARLY were significant [Table 6.13, correlations with sim are also presented for information].

	sim	suc	exec
XLAT	-.08	.11	-.07
XWTLAT	.10	.09	-.10
NCONT	-.15	-.09	.14
NEARLY	-.11	-.16	-.24*
NGOOD	.17	.23*	.30**
NGUESS	.03	-.59	.006

* p < .05 **p < .001

Table 6.13 Correlations between Luria model component scores and *Change of Smoothing Test* variables

A positive correlation between ability on successive synthesis and the number of correct responses suggests that the null hypothesis associated with Hypothesis 2.2 (a) should be rejected. A positive correlation between ability on executive synthesis and the number of correct responses, and a negative correlation between ability on executive synthesis and the number of early responses, suggests that the null hypothesis associated with Hypothesis 2.2 (b) should be rejected. Some 14% of the variance of NGOOD can be accounted for by measures of successive and executive synthesis together.

The absence of any relationships between either successive or executive processing and XLAT and XWTLAT, the variables which measured response latencies, is consistent with the presence of a strong positive relationship between executive processing and NGOOD, the number of correct responses. Ability on executive synthesis includes the ability to make appropriate decisions in response to afferent data. On this task it seems that such decisions are independent of response speed, in contrast to Vernon's evidence for a relationship between intelligence and processing speed (e.g., Vernon & Mori, 1992), and in support of Vernon's detractors (e.g., Longstreth, 1986). As Vernon, Nador and Kantor (1985) explain this relationship in terms of working-memory capacity, which Kirby and Das (1990) conceptualise as strength of successive processing, the absence of any correlation between successive synthesis and XLAT and XWTLAT further suggests that these findings can be interpreted as counter evidence to Vernon's position, unless there exist undetected task-related compromises to a non-random distribution of response latencies. In any case, the variables XLAT and XWTLAT were not considered in further analysis.

Further support for the hypothesised results was found from the MANOVA with planned contrasts with the Luria model partitions of sim3, suc3 and exec3 as independent variables, and the set of NGOOD, NEARLY, NGUESS, and NCONT as dependent variables [Appendix I, Tables 6.14 and 6.15]. The multivariate effects were reported in Table 6.07.

NGOOD	F	df	significance p	effect size η^2
sim3 (1) linear	2.85	1,97	.095	.029
suc3 (1) linear	8.11	1,97	.005	.077
exec3 (1) linear	10.55	1,97	.002	.098

Table 6.14 Univariate effects of Luria model component scores and *Change of Smoothing Test* scores for correct response

NEARLY	F	df	significance p	effect size η^2
sim3 (1) linear	1.80	1,97	.183	.018
suc3 (1) linear	5.42	1,97	.022	.053
exec3 (1) linear	8.25	1,97	.005	.078

Table 6.15 Univariate effects of Luria model component scores and *Change of Smoothing Test* scores for early response

The monotonic decrease in the means for NGOOD across the high-medium-low partitions of suc3 and exec3 [Table 6.16] indicated that subjects with high ability on successive synthesis, and subjects with high ability on executive synthesis, were more successful at the task of correctly perceiving a change in the smoothing coefficient. The monotonic increase in the means for NEARLY across the high-medium-low partitions of exec3 [Table 6.17] indicated that subjects with high ability on executive synthesis were more successful at not making a response before there was any change in the smoothing coefficient.

The linear effect of suc3 (1) was significant for NGOOD supporting the hypothesis that performance on the *Change of Smoothing Test* is dependent on successive synthesis. None of the univariate effects for the curvilinear contrast of suc3 (2) were significant. The univariate effect of exec3 (1) was significant for NGOOD and NEARLY, supporting the hypothesis that performance on the *Change of Smoothing Test* is dependent on executive synthesis. None of the

NGOOD	partitions	exec3	suc3
	high	13.6*	13.6*
	medium	12.6	11.6
	low	9.9*	10.8*

* significant linear effect

Table 6.16 Means of correct response scores on the *Change of Smoothing Test* for high-medium-low partitions of successive and executive synthesis

---

NEARLY	partitions	exec3
	high	7.5*
	medium	10.5
	low	11.2*

* significant linear effect

Table 6.17 Means of early response scores on the *Change of Smoothing Test* for high-medium-low partitions on executive synthesis

---

univariate effects for the curvilinear contrast of exec3 (2) were significant. This indicates that the task of deciding when a pitch structure has changed requires both successive and executive synthesis, with nearly 18% of the task variance explained by differences between high and low information processing abilities.

The responses to the control items were further investigated by computing the mean smoothing coefficient of controls perceived (ACONT1) and the mean smoothing coefficient of controls missed, i.e., falsely responded to (ACONT0). A paired t-test was undertaken on the differences between these means for those subjects who did correctly perceive at least one control, N = 33 [Appendix I]. The difference approached significance (p = .062, two-tailed). The mean smoothing coefficient of controls perceived was less than the mean smoothing coefficient of controls missed [Table 6.18]. This result could indicate that perception of correlation is easier when the stimulus has a higher degree of coherence, which is consistent with Bregman's account of auditory streaming.

---

	controls perceived (no response)	controls missed (response)
mean smoothing coefficient A	3.9	5.3

Table 6.18 Mean smoothing coefficient for *Change of Smoothing Test* control items perceived and missed

---

Hypothesis 2.3: That there will be positive relationships between scores on the MEK and (a) success on the *Scale Estimation of Smoothing Test*, and (b) success on the *Change of Smoothing Test*.

As recommended in Study 1, the MEK1 and MEK5 scores were treated as continuous variables. A one-tailed Pearson correlation between MEK1 and MEK5 was significant (r = .38, p < .001). The two sets of scores were added for a total MEK score (MEKTOT). Three two-tailed correlations between MEK1, MEK5 and MEKTOT and each of the variables reflecting performance on the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* were undertaken [Appendix I]. Correlations were significant for RMSD, NGOOD and NEARLY [Table 6.19]. Again it should be noted that RMSD is an error term, and thus the negative value for r indicates a positive correlation. Also as above, NGOOD, the variable representing the number of correct responses, and NEARLY, the variable representing the number of early responses, best reflected performance on the *Change of Smoothing Test.* Clearly the null hypotheses associated with Hypotheses 2.3 (a) and (b) should be rejected. Abilities in pitch discrimination and pattern recognition were positively correlated with abilities to perceive coherence in auditory structure. This result is possibly in contrast to Cook (1994) and Serafine (1988) who argue that the elements of music cognition need not reflect the formal structure of musical analysis.

	RMSD	NGOOD	NEARLY
MEK1	-.25**	.12	-.18*
MEK5	-.21*	.29**	-.21*
MEKTOT	-.32***	.25**	-.25**

* p < .05   ** p < .01   *** p < .001

Table 6.19 Correlations between MEK scores and measures on the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test*

Hypothesis 2.4: That there will be positive relationships between scores on the MEK and measures of (a) simultaneous synthesis, (b) successive synthesis, and (c) executive synthesis.

This Hypothesis was tested with a series of two-tailed correlations between MEK1, MEK5 and MEKTOT and each of sim, suc and exec [Appendix I]. There were significant correlations for sim, suc and exec [Table 6.20]. This finding supports the general hypothesis that music ability is underpinned by abilities on all three dimensions of information processing. Specifically, the three dimensions of information processing account for over 23% of the variance (Σr^2) of total MEK scores. Clearly the null hypotheses associated with Hypotheses 2.4 (a), (b) and (c) should be rejected.

	sim	suc	exec
MEK1	.16	.27**	.14
MEK5	.25**	.26**	.23*
MEKTOT	.26**	.35***	.22*

* p < .05 ** p < .01 *** p < .001

Table 6.20 Correlations between MEK scores and measures of simultaneous, successive and executive synthesis

This result can be explained by consideration of the processing demands of the MEK test items. In both Part I and Part V most items required recognition of pitch contour, ability in which, as was suggested in Study 1, is dependent on successive synthesis, the predominant dimension of information processing available to young children. 'Visualising' a pitch or melodic pattern for comparison with a graphic representation, as required in the MEK Parts I and V, is a task which depends on simultaneous synthesis. As the ability to direct attention contributes to performance on any tasks of discrimination, executive synthesis is also important for success on these MEK tests.

These findings were further supported by the univariate effects of both the MANOVA using planned contrasts with sim3, suc3 and exec3 as independent variables, and MEK1, MEK5 as dependent variables, and an ANOVA with planned contrasts with the same independent variables and MEKTOT as the dependent variable [Appendix I, Tables 6.07, 6.21, 6.22 and 6.23]. The linear effect of suc3 (1) was significant for MEK1. The interaction effect of suc3 x exec3 (1) was significant for MEK5. The linear effect of exec3 (1) was significant for MEK5. None of the effects for the curvilinear contrasts suc3 (2) or exec3 (2) were significant.

MEK1	F	df	significance p	effect size η^2
suc3 (1) x exec3 (1)	4.46	1,97	.037	.044
sim3 (1) linear	1.76	1,97	.188	.018
suc3 (1) linear	13.61	1,97	<.001	.123
exec3 (1) linear	2.53	1,97	.115	.025

Table 6.21 Univariate effects of Luria model component scores and MEK Part I scores

MEK V	F	df	significance p	effect size η^2
suc3 (1) x exec3 (1)	7.62	1,97	.007	.073
sim3 (1) linear	4.45	1,97	.038	.044
suc3 (1) linear	4.47	1,97	.037	.044
exec3 (1) linear	7.40	1,97	.008	.071

Table 6.22 Univariate effects of Luria model component scores and MEK Part V scores

MEKTOT	F	df	significance p	effect size η^2
suc3 (1) x exec3 (1)	6.52	1,103	.012	.060
sim3 (1) linear	3.22	1,103	.076	.030
suc3 (1) linear	14.64	1,103	<.001	.124
exec3 (1) linear	5.92	1,103	.017	.054

Table 6.23 Univariate effects of Luria model component scores and MEK Part V scores

As the suc3 x exec3 interaction effects were significant, the interaction means of the MEK scores, rather than the main effect means, are reported [Appendix I, Tables 6.24 and 6.25]. The MEK test and total scores are similar for the high and medium levels of successive and executive synthesis, and decrease sharply in the suc3 - low x exec3 - low cell. These results support the interpretation that music perceptual tasks such as pitch discrimination and melodic pattern recognition require the attentional processes of executive synthesis to be informed by successively encoded information such as pitch contour.

MEK1	suc3 - high	suc3 - medium	suc3 - low
exec3 - high	15.18	15.64	15.69
exec3 - medium	15.37	14.00	14.22
exec3 - low	13.90	13.58	11.93

MEK5	suc3 - high	suc3 - medium	suc3 - low
exec3 - high	16.00	16.29	16.23
exec3 - medium	15.95	16.18	15.11
exec3 - low	16.40	16.08	13.29

Table 6.24 Means of MEK test scores for high-medium-low partitions of successive x executive synthesis

MEKTOT	suc3 - high	suc3 - medium	suc3 - low
exec3 - high	31.18	31.93	31.92
exec3 - medium	31.32	30.18	29.33
exec3 - low	30.30	29.67	25.21

Table 6.25 Means of MEK total scores for high-medium-low partitions of successive x executive synthesis

Hypothesis 2.5: That subjects with high degrees of success on the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* will have more music education experience than subjects with less success.

Subjects were grouped on the basis of their music education experience into five categories described in *Section 6.3*. The hypothesis was tested with a 5-group MANOVA with MUSICEXP as the independent variable and RMSD, and the set of NGOOD, NEARLY, NGUESS and NCONT as dependent variables [Appendix I]. None of the effects were significant. As the few subjects in high music-experience categories 4 and 5 may have led to a number of cells with an unacceptably low cell size, the subjects were re-grouped into two categories of school-only and beyond-school music experience (MUSEXP). The analysis was repeated with a 2-way ANOVA with MUSEXP as the independent variable and the same dependent variables. Again, none of the effects reached significance. It seems clear that the null hypothesis associated with Hypothesis 2.4 should not be rejected. This result supports the position that there is a perceptual dimension of musical aptitude which is independent of music education (e.g., Dowling, 1995a; Cuddy and Upitis, 1992; Gordon, 1979, 1993). It could also be noted that the results of the ANOVA with planned contrasts with sim3, suc3 and exec3 as independent variables and MUSICEXP as the dependent variable showed no significant linear or curvilinear effects [Appendix I]. This could indicate that the stability of information processing abilities with age (Das *et al*, 1994) holds in the domain of music.

Hypothesis 2.6: That subjects with high degrees of success on the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* will have higher school achievement levels in (a) mathematics and (b) English than subjects with less success.

The school achievement levels in mathematics and English were each coded on a four-point scale described in *Section 6.3*. Hypothesis 2.6 was tested with a series of two-tailed correlations between MATH and LANG, and each of the sensitivity to autocorrelation structure variables RMSD, NGOOD, NEARLY, NGUESS and NCONT [Appendix I]. The correlations were significant for RMSD and NGOOD [Tables 6.26 and 6.27].

As above, the negative r value for RMSD is an indication of a positive relationship with success on the *Scale Estimation of Smoothing Test*, with some 25% of the variance being accounted for by mathematics performance level, and over 20% of the variance being accounted for by English language performance level. Success on the *Change of Smoothing Test* was again best represented by the number of correct responses, where nearly 18% of the variance was accounted for by mathematics performance level, and 16% of the variance by English language performance level. Clearly the null hypothesis associated with Hypothesis 2.6 (a) should be rejected.

RMSD by	Pearson r	significance p	effect size r^2
MATH	-.50	< .001	.250
LANG	-.45	< .001	.203

Table 6.26 Correlations between school performance and *Scale Estimation of Smoothing Test* scores

NGOOD by	Pearson r	significance p	effect size r^2
MATH	.42	< .001	.176
LANG	.40	< .001	.160

Table 6.27 Correlations between school performance and *Change of Smoothing Test* correct response scores

As the distribution of the MATH and LANG scales were constructed to be non-normal [*Section 6.3*], it was decided to check these results by using the scale ratings as grouping variables in a pair of 4-group MANOVAs for MATH and LANG as the independent variables on RMSD and the set of NGOOD, NEARLY, NGUESS and NCONT as dependent variables [Appendix I]. The multivariate effects for MATH and LANG were significant [Table 6.28]. Some 34% of the variance of RMSD and NGOOD was accounted for by the level of school performance in mathematics and English. The significant univariate effects are reported in Table 6.29, and the means for these effects are reported in Tables 6.30 and 6.31.

The univariate effect for MATH was significant for RMSD. The means for RMSD increase monotonically from MATH (4) (highest) to MATH (1) (lowest) confirming a positive relationship between scale-wise sensitivity to the autocorrelation structures in fBm tone series and performance at school mathematics [Table 6.30]. Scheffe post-hoc tests showed that the differences in mean RMSD scores between MATH (4) and MATH (3), MATH (4) and MATH (2), MATH (4) and MATH (1), and MATH (3) and MATH (1) were significant. Over 27% of the variance of success on the *Scale Estimation of Smoothing Test* can be explained by level of school mathematics performance.

multivariate effect	Wilk's λ	significance p	multivariate effect size
MATH	.594	< .001	.159
LANG	.549	< .001	.181

Table 6.28 Multivariate effects for school performance and sensitivity to autocorrelation tasks

univariate effect	DV	F	df	significance p	effect size η^2
MATH	RMSD	14.28	3,114	<.001	.273
	NGOOD	8.88	3,114	<.001	.189
	NEARLY	3.15	3,114	.028	.077
LANG	RMSD	11.06	3,114	<.001	.225
	NGOOD	7.66	3,114	<.001	.168
	NEARLY	6.82	3,114	<.001	.152
	NGUESS	3.44	3,114	.019	.083

Table 6.29 Univariate effects for school performance and sensitivity to autocorrelation tasks

The univariate effect for MATH was also significant for NGOOD. The mean number of correct responses to the *Change of Smoothing Test* have a bimodal distribution across levels of mathematics performance, from a similar lower level on MATH (1) and MATH (2) to a similar higher level on MATH (3) and MATH (4) [Table 6.30]. This is consistent with a positive relationship between the ability to detect change in autocorrelation structures in fBm tone series and performance at school mathematics. Scheffe post-hoc tests showed that the differences in mean NGOOD scores between MATH (4) and MATH (2), and MATH (3) and MATH (2) were significant. Nearly 19% of the variance of success on the *Change of Smoothing Test* can be explained by level of school mathematics performance.

The univariate effect for MATH was significant for NEARLY. A similar bimodal distribution of means to NGOOD is shown with NEARLY [Table 6.30]. A high level of early responses was made by subjects in MATH (1) and MATH (2) compared with a lower level of early responses made by subjects in MATH (3) and MATH (4). Scheffe post-hoc tests showed that the differences in mean NEARLY scores between MATH (2) and MATH (3) were significant. Consistent with the result above, subjects with high levels of mathematics performance made significantly fewer false responses to unchanging stimuli. Nearly 8% of the variance of avoiding false responses on the *Change of Smoothing Test* can be explained by level of school mathematics performance.

MATH	level 4 (highest)	level 3	level 2	level 1 (lowest)
RMSD	2.49*	3.12*†	3.46*	4.51*†
NGOOD	13.82†	12.15*	8.75†*	8.80
NEARLY	9.05	8.80*	12.65*	10.80

*† Scheffe post-hoc

Table 6.30 Means of *Scale Estimation of Smoothing Test* and *Change of Smoothing Test* scores for school performance in mathematics

Here then is evidence for the much observed relationship between music and mathematics. The considerable effect sizes suggest that this relationship could be one of shared variance in ability, and certainly more than just mere musical interest shown by mathematicians, as suggested by Gardner (1993).

The univariate effect of LANG was significant for RMSD. The means for RMSD increase monotonically from LANG (4) (highest) to LANG (1) (lowest) [Table 6.31], confirming a positive relationship between scale-wise sensitivity to the autocorrelation structures in fBm tone series and performance at school English. Scheffe post-hoc tests showed that the differences in mean RMSD scores between LANG (4) and LANG (3), LANG (4) and LANG (2), LANG (4) and LANG (1), and LANG (3) and LANG (1) were significant. Over 22% of the variance of success on the *Scale Estimation of Smoothing Test* can be explained by level of school English language performance.

The univariate effects for LANG was significant for NGOOD. The mean number of correct responses to the *Change of Smoothing Test* increases monotonically across levels of English language performance from LANG (1) to LANG (4) [Table 6.31]. This is consistent with a positive relationship between success at perception of changes in the autocorrelation structures in fBm tone series and performance at school English. Scheffe post-hoc tests showed that the differences in mean NGOOD scores between LANG (4) and LANG (3), and LANG (4) and LANG (2) were significant. Nearly 17% of the variance of success on the *Change of Smoothing Test* can be explained by level of school English performance.

The univariate effect for LANG was significant for NEARLY. There is a curvilinear distribution of NEARLY means across levels of LANG [Table 6.31]. Whereas a low level of early responses was made by subjects in LANG (4) and LANG (3), the highest number of early responses was made by LANG (2) rather than LANG (1). Scheffe post-hoc tests showed that the differences in mean NEARLY scores between LANG (2) and LANG (3), and LANG (2) and LANG (4) were significant. Subjects with high levels of English language performance made significantly fewer false responses to unchanging stimuli. Over 15% of the variance of avoiding false responses on the *Change of Smoothing Test* can be explained by level of school English language performance.

The univariate effect for LANG was significant on NGUESS. The distribution pattern of means is not readily interpretable [Appendix I]. This effect was probably due to a spurious low mean for LANG (2).

In addition to a relationship between music perception and mathematics, these findings show a relationship between language facility and music aptitude, which is also commonly observed by music teachers but is less publicised than the relationship with mathematics. These relationships can be explained by recourse to the Luria model. Success in primary school mathematics and language is accounted for by abilities on successive synthesis, but with increasing dependence on simultaneous and executive synthesis in the senior years (e.g., Crawford, 1986; Hunt, Randhawa & Fitzgerald, 1976; Ransley, 1981; Try, 1984; Tulloch, 1981).

LANG	level 4 (highest)	level 3	level 2	level 1 (lowest)
RMSD	2.63†	3.18*†	3.36†	4.51*†
NGOOD	13.82*	11.25†	9.44*†	8.80
NEARLY	8.80†	8.70*	14.6*†	10.8

*† Scheffe post-hoc

Table 6.31 Means of *Scale Estimation of Smoothing Test* and *Change of Smoothing Test* scores for school performance in English language

Consistent results were found in the univariate effects associated with the MANOVA using planned contrasts, with sim3, suc3 and exec3 as dependent variables, and MATH and LANG as dependent variables [Appendix I, Tables 6.07 and 6.32]. The linear effects of suc3 (1) were significant for MATH and LANG . None of the univariate effects for the curvilinear contrasts of suc3 (2) were significant. The univariate effects of exec3 (1) were significant for MATH and

	DV	F	df	significance p	effect size η^2
sim3 (1) linear	MATH	3.01	1,97	.086	.030
	LANG	4.10	1,97	.046	.041
suc3 (1) linear	MATH	14.33	1,97	<.001	.129
	LANG	27.26	1,97	<.001	.219
exec3 (1) linear	MATH	20.01	1,97	<.001	.171
	LANG	12.59	1,97	<.001	.115

Table 6.32 Univariate effects of Luria model component scores and school performance

LANG. None of the univariate effects for the curvilinear contrasts of exec3 (2) were significant. These results indicate that academic success at mathematics and language studies in primary school requires successive and executive synthesis, with around 30% of the task variance on each of MATH and LANG explained by differences between high and low information processing abilities.

Since the sensitivity to autocorrelation structure tasks are likewise dependent on simultaneous, successive and executive syntheses, particularly the latter two, these findings are consistent with the interpretation above that the strong relationships between school performance and music perceptual abilities are underpinned by common abilities on dimensions of information processing. This interplay between these dimensions of information processing could be the common or underlying factor noted by Sergeant and Thatcher (1974) [Chapter 3] to explain the high correlations they obtained from their re-analysis of data on relationships between musical ability and general intelligence .

A summary of the findings of *Section 6.6* with respect to the research Hypotheses 2.1 to 2.7 is provided in Table 6.33.

Research Hypotheses supported	**Research Hypotheses rejected**
2.1 (a) (b), 2.2 (a) (b), 2.3 (a) (b), 2.4 (a) (b) (c), 2.6 (a) (b)	2.5

Table 6.33 Summary of Study 2 hypothesis testing

6.7 General discussion

The results show that the cognitive processing involved in music perception can be explained in terms of the three information processing dimensions of the operationalised Luria model. The results from the analysis of contrasts between the extreme and curvilinear effects of simultaneous, successive and executive syntheses point to the importance of these dimensions, particularly successive and executive syntheses, for success in music perception and other tasks.

Testing Hypothesis 2.4 (a), (b) and (c) addressed the fourth research question: Do musically able children have higher measures on any of the three Luria model dimensions? That music information processing is undertaken in a similar manner to other modes of information processing is obviously a more parsimonious model for music cognition than models which posit specific music perceivers and processors. Gardner's construct of a separate music intelligence may be operationally convenient on occasions, but these results suggest that separate intelligences are underscored by a few basic cognitive functions. It would be difficult to understand the interaction of various intelligences in real life situations if there were no common functionings. The sixth research question concerning the relationship between measured music ability and school achievement was tested with Hypothesis 2.6. The significant correlation found here between abilities in music perception and mathematics performance supports what teachers have long observed. The significant correlation with language performance is also consistent with teacher observations, and supports the findings of Karma (1982, 1985), and Hassler, Birbaumer and Feil (1985, 1987) who found a correlation between musical aptitude and verbal reasoning with 10 year old subjects.

Such a result would be expected from the cognitive development of simultaneous and executive syntheses in children during their primary schooling. The fifth research question concerning the relationship between measured music ability and music experience was tested with Hypothesis 2.5. When cognitive development takes place within a musically stimulating and nurturing environment, clearly the child will be musically advantaged (e.g., Creider, 1989; Cuddy & Upitis, 1992; Davidson, 1993; Freeman, 1974). However, sensitivity to differences in autocorrelation structure was found to be independent of musical experience, age and school Year. In so much as perceptual ability is a contributor to a more general musical aptitude, the results of Study 2 support the argument that not all of the variance in musical ability can be accounted for by early environmental influences (e.g., Beament, 1977; Boyle, 1992; Davies, 1994; Garner, 1993; Gordon, 1993; Radford, 1994; Torff & Winner, 1994). On the other hand, discrimination of pitch was found to be related to age and school Year, suggesting that cognitive maturity and general non-educational music experience are also important contributors to levels of music ability (Serafine, 1988; Sloboda *et al*, 1994a, 1994b; Storr, 1992).

It would seem reasonable, then, in an educational context, to suggest that such innate ability could act as a potential for learning (Gardner, 1993; McPherson, 1995), or as a ceiling which might be reached depending on circumstances (Gordon, 1979). This position is also consistent with Gagne's more general model of actualisation (Gross, 1993). On the other hand, the argument that everyone is musical (Sloboda *et al*, 1994a) can also be supported to the extent that music perceptual demands are underpinned by information processing dimensions common to all cognitive tasks. Serafine (1988) found that pitch discrimination was uncorrelated with success on tasks of successive and simultaneous temporal processing. She argues that such temporal processes are a part of normal human cognitive development, and as such, are independent of exposure to music education. "The conclusion seems inescapable that, at the least, formal instrumental training is neither necessary nor sufficient for the development of the generic processes. Rather ... the more potent factors may be general cognitive growth and normal, everyday musical experience" (p. 229).

The Luria model has proved to be more robust in this regard than other models of intelligence. All three information processing dimensions, simultaneous, successive and executive synthesis, seem to be required for musical processing, their relative importance depending on the demands of the task, and the abilities of the listener/performer/composer. It must be emphasised that music is not one-dimensional, and like any real-world task, requires a coordinated cognitive response, as Luria emphatically stated.

The roles of the three Luria model information processing dimensions in music can be seen in their different usage in the two sensitivity to autocorrelation structure tasks. The first research question asked whether tone series with different autocorrelation structures are distinguishable from one another. This was implicitly tested by Hypotheses 2.1 (a) and (b), and 2.2 (a) and (b). The results showed that tone series with different autocorrelation structures, as generated by the application of different smoothing coefficients to random tones, are distinguishable from one another. The third research question followed: Within the Luria model, what information processing abilities are important for perceptual sensitivity to differences in autocorrelation structures? This was addressed directly with Hypothesis 2.1 (a) and (b), and 2.2 (a) and (b) . There is clear evidence that abilities to make such discriminations are underpinned by abilities on the three information processing dimensions of the Luria model. Two perceptual characteristics of an auditory stimulus are its contour shape and pitch range (Cohen, Trehub & Thorpe, 1989; Dowling, 1990). In Study 1 it was suggested that variability in contour is encoded by successive synthesis, and pitch by simultaneous synthesis. The predominance of positive relationships between performance on the sensitivity to autocorrelation structure tasks and abilities on successive synthesis in this study suggests that subjects were utilising variability in contour profile rather than pitch range to discriminate between the different tone series. As with the results in Study 1 concerning the perception of 'fractal' music, the degree of perceived strength of autocorrelation structure could be a function of the data length of the fBm tone series (Tong, 1980) which, as was argued in Chapter 5, would be determined by the capacity of a subject's working memory, i.e., by an individual's ability on successive processing (Kirby & Das, 1990). Hantz and Crummer (1988) argue that performance on tasks such as the *Change of Smoothing Test*, where subjects are "able to respond 'different' at the point of the first noticeable difference" (p. 112), is governed by the continual updating of working memory.

Such an interpretation calls into question the model of bipartite perception of complex auditory contours by Gilden *et al* (1993) [Chapter 4]. Highly correlated tone series with low smoothing coefficients have high signal-to-noise ratios. Their structure under a bivariate model should thus be more readily perceived, since a highly correlated structure is mostly signal. Conversely, the structure of tone series with high smoothing coefficients should be more difficult to perceive since the signal will be embedded in the noise. This was not the case in the trial of the *Scale Estimation of Smoothing Test,* where item performance was independent of position on the scale, or with the *Change of Smoothing Test*, where the response latency, weighted by the smoothing coefficient, did not relate to success at perceiving the change. In support of Gilden *et al*, however, there was some evidence that tone series with lower smoothing coefficients were more readily perceivable as controls, consistent with findings from earlier studies into children's perception of tone sequence complexity (McMullen, 1974; McMullen & Arnold, 1976). This matter was taken up in the third study.

In Study 2, the focus was on the relationships between the Luria model dimensions and the perception of autocorrelation structure. Each sensitivity task had characteristic requirements for information processing. Individual differences in performance on these tasks arise from the processing bottlenecks caused by limitations in ability on a particular processing dimension. The information processing demands of the *Scale Estimation of Smoothing Test,* to judge the relative level of coherence of a tone series, make similar calls on the use of simultaneous synthesis as in establishing a tonal centre (Leman, 1992), evaluating a pitch hierarchy (Krumhansl, 1990), or audiating the tonic in an ambiguous keyality (Walters, 1989). The developing ability on simultaneous synthesis of the growing child accounts for his/her increasing ability to flesh out the contours of melodies with appropriate pitches (Dowling, 1994a, 1994b). Limited ability on simultaneous synthesis compromises performance on tasks such as the *Scale Estimation of Smoothing Test.*

The <u>second research question</u>, which was concerned with the sensitivity of musically able subjects to differences in autocorrelation structures, was tested with Hypothesis 2.3 (a) and (b). The results support previous research that musical ability positively affects discrimination performance in atonal contexts (Krumhansl *et al*, 1987). Musically able subjects may bring both a more stable memory for music (Crummer *et al*, 1988; Tsuzaki, 1991) and a superior ability to detect sequential coherence (Bartlett, 1984; Billingsley & Rotenberg, 1982) to these tasks.

The determination of change in the *Change of Smoothing Test* requires successive synthesis to process the information in a similar manner to determining the occurrence of an harmonic modulation, or to noting the phrase boundaries while listening to music unfold in time (Serafine, 1988). Minsky (1988) argues that the mental construction of appropriate phrase boundaries in language is an essential task to comprehension; a string of words has no meaning without a divisional structure to indicate which words should cluster together. So it is with music. The strong relationships found here between abilities on perception of music structure and school performance in English suggests that some subjects have a high common ability to construct appropriate phrase boundaries in both domains.

The determination of change in the *Change of Smoothing Test* also requires executive synthesis in the creation of expectancy as the basis of an evaluation of appropriateness of continuity (Balch, 1981). Deciding that there exists a change which warrants a response as an all-or-nothing event, highlights well the coarse-grained nature of human perception. It is a higher-order or top-down cognitive activity which must be based on the organised input from a lower-order or bottom-up process. Individuals with high ability on executive synthesis are better at this task than their peers. This result is also consistent with Bregman's model of primitive or bottom-up and schema-based or top-down organisation of auditory input. Consequently, the model outlined in Chapter 4, where autocorrelation is a means by which musical information is initially encoded successively for chunking into meaningful units for musical interpretation, is now advanced with more confidence.

The roles of the Luria model information processing dimensions with regard to the MEK tests were more clearly delineated in this study than in Study 1. Two improvements in experimental sensitivity seemed responsible: the use of computer-adaptive operationalisation of the Luria model, and the use of MEK scores as continuous variables. The MEK was designed for classroom music teachers to determine which students at the beginning of secondary school (Year 7) had already reached a standard of musical ability that might be expected at the completion of Year 10. It was assumed in designing the instrument that aptitude, home environment, primary school music program, and individual motivation, would all contribute to the perceived ability level (Bryce, 1979). This is consistent with the literature (e.g., Beament, 1977; Davidson, 1994; Gordon, 1993; Taylor, 1973), that ability at pitch perception is acquired through learning. This may explain the positive relationship between MEK Part I scores and subject age. As a criterion measure, however, the focus is on achievement rather than aptitude or potential. As a normative measure, the relationships between task performance and other abilities can better be examined. Performance on Part I *Pitch Discrimination* is underpinned by the ability to employ successive synthesis to encode pitch contour. Performance on Part V *Pattern Recognition* requires attentional processes, possibly to coordinate the chunking of serial tones into a holistic pattern. That is, the levels of music abilities acquired through learning are determined and shaped by the information processing abilities of the individual. This process is non-linear to the limited extent that information processing abilities are enhanced through experience [Chapter 2].

This interpretation points to some of the strengths and limitations of this study. A particular strength of the study is that the Luria model is multidimensional. Karma (1985) notes that "several cognitive operations [are] involved in structuring sequences of auditive material: forming expectations, recognising, structuring according to Gestalts, changing expectations, timing, and analysing the internal structures of strong Gestalts" (p. 11). As the stimulus tone series in these sensitivity to autocorrelation structure tasks lacked conventional keyality and harmonisation, whatever Gestalts could be constructed would presumably be novel for all subjects. This was seen as another strength of the study, in that the effects of music experience would be controlled. The failure to find any relationship between sensitivity to autocorrelation structure and music experience confirms this expectation. Again this supports Bregman's model where bottom-up streaming correlates acoustic clues but pitch trajectory as such does not contribute to schema-based segregation. Furthermore, an increase in tempo enhances primitive streaming but worsens perceptual streaming. It is interesting to note how many of the weaker subjects on the *Scale Estimation of Smoothing Test* labelled the A = 1 example as "fast".

On the other hand, a clear limitation to the generalisability of these results for music cognition is that fBm tone series are not music. Although stochastic computer-generated music enjoys a small but growing share of the contemporary repertoire, it could not be claimed that the tone series of the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* were examples of real music to the ears of 10 and 11 year old subjects. As such, the use of fBm tone series opens the argument that performance on these tests does not necessarily predict performance in real music contexts. Such an argument against clinical testing is hardly novel (e.g. Brown, 1988; Unyk, 1990). This limitation was partly offset by the use of tones with conventional pitch, and, more fundamentally, by using tone series which exhibit self similarity in form, as was purportedly demonstrated for real music by Hsu (1993). In *Section 6.2* it was noted that Turner (1992) gives $A \approx 4.5$ as the smoothing coefficient required to produce tone series which have a spectral density function coefficient $\beta = 1$, the value Voss argues is characteristic of real music. An investigation of possible relationships between perceptual sensitivity and strength of autocorrelation structure was included in the third study.

6.8 Recommendations for a third Study

Whereas the results from Study 2 support the thesis that individual differences in music perception can be explained in terms of the information processing dimensions of simultaneous, successive and executive synthesis, there were several contingencies to such a claim which might profitably be addressed in a further study.

The results of the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* were very similar to the results from the instrument trials, suggesting that reliability is satisfactory. Also, the near normal distribution of the relevant variables is consistent with a valid measurement of individual differences. Nevertheless, a replication study with these new instruments would seem advisable. As the question of generalisability is not so readily resolved, there could be an advantage if the subjects of a third study were selected on the basis of an already demonstrated high level of musical ability. Certainly the suggestion that the *Scale Estimation of Smoothing Test* could act as a marker test for simultaneous synthesis in the music domain would need exploration with high music ability subjects.

Such a study could focus on the important role for executive synthesis in the establishment of expectancy and attention (Meyer, 1970 ; Pogonowski, 1985; Webster & Richardson, 1994) [Chapter 3]. For musically talented children it might be conjectured that the self-regulatory function of executive synthesis (Crawford, 1986) provides positive emotional feedback for motivation to maintain a disciplined regimen of practice and study. The interactive effects of sim x exec and suc x exec on the perceptual variables in Study 1 and Study 2 respectively suggest that the role of executive synthesis bears further investigation.

It would seem that these concerns could be addressed in a third study involving subjects with demonstrable music ability, either through performance at professional level while still very young, or selection into advanced music training programs, or success at music exams at a superior level to the norm, or who show a rate of learning music well above that of their peers. This third study should employ the *Scale Estimation of Smoothing Test* and the *Change of Smoothing Test* together with the computer-adaptive Luria model battery. As a test of the validity of the application of the Luria model to music information processing, it could be hypothesised that individuals with such high music ability would have high measures on all three of simultaneous, successive and executive syntheses. To test the model proposed in Chapter 4, it could also be hypothesised that highly musical subjects would show significantly superior performance to that of the large sample from Study 2 on both the perception of autocorrelation structure tests. To that end, the best variables on which to compare performances would be RMSD for the *Scale Estimation of Smoothing Test* and NGOOD for the *Change of Smoothing Latency Test.*

Since Bregman argues that good musical composition facilitates auditory functioning, an investigation, as suggested above, into possible relationships between the perception and strength of autocorrelation structure might inform the argument over the autocorrelation structure, and thus the power spectrum of real music: $\beta = 1$ (Voss & Clarke, 1975, 1978; Hsu, 1993) or $\beta < 1$ (Nettheim, 1992; Boon & Decroly, 1995; Gregson, 1994). Such an endeavour would be supported by using subjects with well developed, stable and accurate schema of conventional musical structure. Subjects selected on this criterion would also be suitable for testing Gilden *et al*'s bipartite model of perception of auditory contour.

Chapter Summary

This chapter presented the results of the second study with upper primary school children as subjects. Two original instruments for testing the perceptual penetrability of differences in autocorrelation structure were developed, one hypothesised to require simultaneous synthesis, the other successive synthesis. The results supported these hypothesised relationships. Other aspects of musicality, such as innate musical potential and the contribution to performance of musical experience, were also explained within the Luria model. A significant relationship was found between success on tasks of perception of autocorrrelation structure and mathematics and English performance levels, which again was explained within the Luria model. Recommendations for a third study with highly musical subjects were outlined.